



Hendry, A., & Holmboe, K. (2021). Development and validation of the Early Executive Functions Questionnaire: A carer-administered measure of Executive Functions suitable for 9- to 30-month-olds. *Infancy*. <https://doi.org/10.1111/infa.12431>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1111/infa.12431](https://doi.org/10.1111/infa.12431)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1111/infa.12431>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Development and validation of the Early Executive Functions Questionnaire: A carer-administered measure of Executive Functions suitable for 9- to 30-month-olds

Alexandra Hendry¹  | Karla Holmboe^{1,2} 

¹Department of Experimental Psychology, University of Oxford, Oxford, UK

²School of Psychological Science, University of Bristol, Bristol, UK

Correspondence

Alexandra Hendry, Department of Experimental Psychology, University of Oxford, Oxford, UK.

Email: alexandra.hendry@psy.ox.ac.uk

Funding information

The authors declare no conflicts of interest with regard to the funding source for this study.

Abstract

Executive functions (EFs) enable us to control our attention and behavior in order to set and work toward goals. Strong EF skills are linked to better academic performance, and greater health, wealth, and happiness in later life. Research into EF development has been hampered by a lack of scalable measures suitable for infancy through to toddlerhood. The 31-item Early Executive Functions Questionnaire (EEFQ) complements temperament measures by targeting cognitive and regulatory capabilities. Exploratory Factor Analysis ($n = 486$ 8- to 30-month-olds) and Confirmatory Factor Analysis ($n = 317$ 9- to 30-month-olds) indicate Inhibitory Control, Flexibility, and Working Memory items load onto a common “Cognitive Executive Function (CEF)” factor, while Regulation items do not. The CEF factor shows strong factorial measurement invariance for sex, and partial strong factorial measurement invariance for age. CEF and Regulation scores show limited floor and ceiling effects, good internal consistency, short-term stability, and convergent validity with carer-report measures of attentional control. The EEFQ is sensitive to developmental change. Results indicate that the widely

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Infancy* published by Wiley Periodicals LLC on behalf of International Congress of Infant Studies

overlooked period between late infancy and early toddlerhood may be a sensitive period for EF development. The low-resource demands of the EEFQ afford the possibility to study emergent EFs at scale; opening up new opportunities in basic developmental and intervention research.

1 | INTRODUCTION

Executive functions are the skills required for top-down control of attention and behavior. These skills enable us to resist acting on impulse, adjust our actions during a changing situation, and work toward goals. Early EF skills have been implicated in the development of a range of social and cognitive skills, including theory of mind, and language (Carlson et al., 2002, 2015; Hughes, 1998; Weiland et al., 2014). Early EF skills emerge during the second half of the first year, develop rapidly within the first 3 years of life, and show some predictive associations with later behavior (Hendry et al., 2016; Mulder et al., 2014).

For many years, the emphasis in infant EF research has been on identifying universal patterns of EF development, using tightly controlled homogenous samples to understand the impact of manipulations of experimental task conditions on group performance. The extensive body of work investigating the effects of manipulations of the A-not-B object retrieval task (considered a measure of response inhibition) is a prime example of this (Marcovitch & Zelazo, 1999). More recently, attention has increasingly focused on the study of individual differences in EFs as a means of delineating developmental mechanisms, and predicting influences on—and consequences of—early individual variation (Hendry et al., 2016; Hughes et al., 2020; McHarg et al., 2020; Pérez-Edgar et al., 2020). When embedded in longitudinal designs, such work has potential to illuminate the impact of environmental factors on EF development and to inform intervention design for populations showing, or at risk for, EF difficulties. However, this approach requires large samples to be adequately powered to detect the small effects characteristic of infant individual differences (Pérez-Edgar et al., 2020).

One cost-effective approach to achieving the large sample sizes required for individual differences research is to use questionnaires. Questionnaires can be administered remotely, at low cost and with low demands on participants. An advantage of using parent-report questionnaires to study early development is that primary care tends to be the responsibility of a small number of adults—at least in Western societies—so that a primary carer (frequently a parent, such that the term parent is often used, as here, as short-hand for any primary carer) has the opportunity to observe their child both in a range of contexts and over a sustained period. Parent report thus provides an insight into child behavior that is both broad and deep (Rothbart & Mauro, 1990).

Further, parent report may be more sensitive to different, albeit complementary, aspects of EF to lab-based performance measures. Toplak et al. (2013) have argued that existing parent-report measures capture individual differences in success in goal pursuit, whereas lab measures tend to be more sensitive to the efficiency of cognitive abilities. Indeed, Nelson et al. (2016) have demonstrated that among 3- to 4-year-olds, but not 4- to 5-year-olds, individual differences in performance on putative EF tasks—which in that study included extensive language demands, such as naming colors, shapes, animals and objects, and following fairly complex instructions—are



influenced primarily by variation in broader cognitive abilities such as processing of sensory inputs, motor control, and language ability.

Criticisms of early temperament questionnaires focused on the potential bias of parent report, for example, parents' mental health may influence their ratings of the infant (Vaughn et al., 1981). However, modern instruments address this by refraining from asking emotive and comparative questions about the infant and instead focusing on observed behaviors and their frequency within the last 1–2 weeks (Rothbart, 2011).

There are a number of EF questionnaires suitable for children and adults—the most commonly used being the Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2002) and the BRIEF-Preschool version (BRIEF-P; Gioia et al., 2002) for 2- to 6-year-olds. The BRIEF-P shows promise of sensitivity to different manifestations of EF difficulties in the context of clinical conditions such as autism and ADHD (Ezpeleta & Granero, 2015; Sherman & Brooks, 2010; Skogan, Zeiner, et al., 2015; Smithson et al., 2013), small-to-moderate associations with concurrent and later academic skills (C. A. C. Clark et al., 2010; Spiegel et al., 2017) and possible associations with variation in brain structure (Ghassabian et al., 2013). More recently, the Ratings of Everyday Executive Functioning (REEF) (Nilsen et al., 2017) has been introduced as a measure of preschoolers' EF abilities in day-to-day life. However, neither the BRIEF-P, nor the REEF are validated for use with children under 2 years of age, leaving an important gap in our ability to measure EFs via carer report as they first emerge. Indeed, investigation into change and stability in very early EF development has been hampered by an absence of measures suitable for use across the infant-to-toddler period (Hendry et al., 2016; Petersen et al., 2016).

The parent-report questionnaires most relevant to early EF which are currently available for infants and toddlers are the Infant Behavior Questionnaire-Revised (IBQ-R) for 3- to 12-month-olds (Gartstein & Rothbart, 2003), and the Early Childhood Behavior Questionnaire (ECBQ) for 18- to 36-month-olds (Putnam et al., 2006). These questionnaires were developed to assess a range of temperament traits, which were subsequently organized into the broad dimensions of Surgency/Extraversion, Negative Affectivity and Orienting/Regulation or Effortful Control (Putnam et al., 2006; Rothbart, 2011). Broadly, the evidence indicates that multiple aspects of temperament play a role in the development of Executive Functions. For example, at age 10 months, Surgency shows a positive concurrent association with behavioral measures of sustained attention and a positive predictive association with A-not-B performance at age 18 months (Frick et al., 2018), while Effortful Control has been found to associate with aspects of EF from 2.5 years of age (Gerardi-Caulton, 2000; Rothbart et al., 2003).

Although some researchers report and interpret Orienting/Regulation and Effortful Control factor scores as if they are synonymous with EF, these measures were intended to capture the regulatory dimension of temperament and were not designed as indices of cognitive function (Gartstein & Rothbart, 2003; Putnam et al., 2006). Indeed, although some of the contributing scales to these factors—notably Attentional Focusing and Attention Shifting—index attentional control, which is considered foundational to EF development (Hendry et al., 2016), others—for example, Cuddliness and Low Intensity Pleasure—are only tangentially related to EF. Further, many constructs generally considered to be core components of EF are either missing entirely (e.g., working memory and cognitive flexibility) or are only partially represented. For example the Inhibitory Control scale of the ECBQ only taps response-to-prohibition type behaviors.

Moreover, considerable differences between the scales included in the IBQ-R and the ECBQ hinder longitudinal measurement of these skills from infancy. For example, the Inhibitory Control scale of the ECBQ is not included in the IBQ-R, while the Duration of Orienting scale of the IBQ-R captures different behaviors to the Attentional Focusing scale of the ECBQ and may

conflate strengths in sustaining attention with difficulties with disengagement, particularly early in infancy (Hendry et al., 2016). This may explain why, despite the well-established links between attentional control and EF (Anderson, 2002; Hendry et al., 2019; Petersen & Posner, 2012) a recent well-powered study failed to find an association from 4-month Duration of Orienting scores to behavioral measures of EF at 14 months (Devine et al., 2019). Therefore, we aimed to develop and validate a new measure of EF—the Early Executive Functions Questionnaire (EEFQ)—which would fill the measurement gap in terms of carer report of early executive functioning in infancy and toddlerhood.

There is not yet a clear consensus on the structure of very early EF—in part due to the shortage of suitable measures which motivated the development of the EEFQ. Some previous work has indicated that a unitary latent EF construct best describes preschoolers' performance on EF batteries (Hughes et al., 2009; Nelson et al., 2016; Senn et al., 2004; Wiebe et al., 2008, 2011; Willoughby et al., 2016; Willoughby et al., 2012)—but other studies have detected dissociable EF factors in children aged 2 and 3 (Bernier et al., 2010, 2012; Garon et al., 2014, 2016; Mulder et al., 2014; Skogan, Egeland, et al., 2015). Some researchers have made a distinction between “cool” and “hot” EF (Mulder et al., 2014; Zelazo & Carlson, 2012), where cool EF is engaged in tasks involving abstract problems such as the selective application of a rule, and where no extrinsic motivator for performance is included, while hot EF is engaged when suppressing an emotionally charged response to a desirable object.

To enable us to collect data which takes into account this debate about the structure of EF, we organized the scale development process of the EEFQ (outlined in Study 1) around the more granular domains which have been adopted within the early EF literature. In turn, this literature has been influenced by Miyake and colleagues' work demonstrating that EFs show both unity and diversity among young adults (Miyake & Friedman, 2012; Miyake et al., 2000), and by Posner's model of attentional control which argues for the importance of executive attention in exerting top-down control, monitoring conflict and maintaining self-regulation (Petersen & Posner, 2012). Our 6 core domains were as follows:

1. **Inhibitory control.** In the Miyake and Friedman (2012) model, performance on inhibitory control tasks is driven by a Common EF latent variable.
2. **Regulation.** Although omitted from most laboratory measures, emotion regulation is an important dimension of observer report measures of EF, reflecting its influence on day-to-day functioning (Isquith et al., 2004; Skogan, Egeland, et al., 2015; Spiegel et al., 2017);
3. **Working memory.** Performance on working memory tasks is most strongly linked to the Updating latent variable in the Miyake and Friedman (2012) model.
4. **Flexibility.** Performance on cognitive flexibility tasks is most strongly linked to the Shifting latent variable in the Miyake and Friedman (2012) model.
5. **Attentional control.** Although not included in the Miyake and Friedman (2012) model, other work by Friedman and colleagues indicates that childhood attention problems, such as difficulties with maintaining attention, are negatively associated with later EFs, particularly inhibitory control (Friedman et al., 2007). In infancy, attentional control (most commonly maintaining attention on a target) shows moderate and predictive associations with aspects of inhibitory control and cognitive flexibility from around 9 months (Hendry et al., 2016).
6. **Persistence, planning, and problem-solving.** As higher-order constructs, these are not generally represented as components in theoretical models but are a distinct dimension of observer report measures of EF (Hughes et al., 2009; Isquith et al., 2004; Senn et al., 2004; Skogan, Egeland, et al., 2015; Spiegel et al., 2017).



As shown in Table S2.1, even at the concept-mapping stage we identified areas of considerable overlap between domains, such that as well as internal consistency within scales (i.e., high correlations between items mapped to the same domain), we expected at least some positive association between scales. We evaluate this overlap empirically in Study 2.

1.1 | Study design

Below we set out a series of studies which, in combination, had 4 aims:

- Aim 1: Test whether the EEFQ can be used to measure EF in infants as young as 9 months as demonstrated through: low rates of missing items; adequate internal consistency on a priori theory-driven scales; sensitivity to developmental change; good short-term stability.
- Aim 2: Investigate the factor structure of EEFQ data using a combination of exploratory and confirmatory approaches across multiple samples.
- Aim 3: Establish convergent validity of the EEFQ with existing measures of attentional control, and investigate convergent and discriminant associations with broader aspects of temperament.
- Aim 4: Present empirical data relating to developmental change and stability in parent-reported EF during infancy and toddlerhood.

2 | STUDY 1: DEVELOPMENT AND REFINEMENT OF THE EEFQ

2.1 | Scale development

Following a review of the literature summarized above, we identified 52 target constructs (facets of EF) mapped to 6 core domains: Inhibitory control, Flexibility, Working memory, Regulation, Attentional control, and Persistence, planning and problem-solving; see Supplementary Materials (SM) 2 for further detail. We drafted 62 items to measure these target constructs (see Table S2.1) The initial item pool was deliberately over-inclusive; designed to sample all possible aspects of the target constructs (Clark & Watson, 2019). To minimize reporter bias, items were framed to ask about recently occurring events and concrete infant behaviors rather than requiring parents to make abstract or comparative judgments (Nilsen et al., 2017; Rothbart & Goldsmith, 1985).

As EF is implicated in almost all aspects of day-to-day life, we aimed to link items to a wide range of activities that an infant might be expected to engage in, including toy play, instruction-following and exploring the home. Where we expected a behavior or skill to be relatively infrequently used in day-to-day life and therefore difficult for parents to report accurately, we identified games for parents to play with their children which would elicit that particular behavior. For example, updating a mental representation of a hidden object frequently appeared as a target construct in the literature but might be difficult for a parent to judge. We therefore outlined a simplified version of the A-not-B task (Diamond, 1985) for parents to administer at home. Other games were used to provide a standardized context for particular EF skills that may vary considerably depending on the situation—such as the child's ability to withhold a response when requested. Eight games were devised for the initial item pool (see Table S2.1). Parents were given detailed instructions on how to administer the games (see SM1), and given the option to play and

score them now, or return to them at the end of the questionnaire. Items were designed to minimize language demands where possible by focusing on infants' spontaneous behavior during play. For the 6 final items (3 of which were games) that did involve explicit language input, items were worded to allow parents to use gesture and demonstration, and/or adapt phrases to their native language or baby-talk.

Recognizing that many researchers will want to collect parent report of EF alongside established measures of control of attention (i.e., the Attention Focusing and Attention Shifting scales of the ECBQ) and/or broader aspects of temperament, we used the same 7-item Likert response scale ranging from Never to Always as the ECBQ (see SM3). Not only does this approach maximize comparability of the EEFQ and ECBQ attention scales, it also means that for studies where the EEFQ is combined with IBQ-R or ECBQ items, respondents do not have to adjust to different rating scales.

The initial item pool was iteratively refined for face validity, clarity, and age-appropriateness via expert review and user testing. Specifically, we asked a panel of developmental psychologists at University of Oxford to comment on whether each item tapped the target construct, was worded unambiguously, and was appropriate for 9- to 30-month-olds. We also conducted semi-structured interviews with 8 parents of 10-month-olds (as EF behaviors are most challenging to measure at the bottom end of the age range) in which we checked that parents: understood the item as currently worded; considered the item to be appropriate to their child's developmental stage and day-to-day life; and were able to give a rating for the item for their child. Items were refined or discarded as appropriate, so that the draft EEFQ was reduced to 44 items (7 game-based), mapped a priori to 4 scales: "Inhibitory Control," "Flexibility," "Working Memory" and "Regulation." Separate "Attentional Control" and "Persistence, Planning and Problem-solving" scales were not retained as items explicitly targeting planning were deemed by respondents to be too challenging, and all other items could be mapped to either Inhibitory Control, Flexibility or Working Memory.

2.2 | Scale refinement

2.2.1 | Participants

Parents or guardians of 9- to 30-month-old children were recruited via email from the Oxford University BabyLab volunteer database between November 9, 2018, and December 23, 2018. This study, and the studies described below, was 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 involving human subjects in this study were approved by the Medical Sciences Inter-Divisional Research Ethics Committee (IDREC), reference R57972/RE002, at the University of Oxford.

After quality checks detailed in SM4.1.1, 65 participants contributed data sufficient to compute all 4 EEFQ scales, 3 contributed valid data for 3 EEFQ scales, and a further 3 contributed valid data for 2 EEFQ scales. The ages and sex of the children described are shown in Table 1. Ninety-five percent of respondents were the child's mother, and 5% were the father.

Parents completed the draft EEFQ online via a unique link sent by email. Reverse scoring of relevant items was carried out prior to analysis. Internal consistency was evaluated using Cronbach's alpha.

2.3 | Results and discussion

2.3.1 | Item reduction

To refine the EEFAQ, items were removed if—across all age-groups—they:

- were frequently (more than 15% of responses) reported as not applicable: 2 items
- showed substantial ceiling effects (more than 50% of responses scoring the maximum): 3 items (1 game-based)
- showed substantial floor effects (more than 50% of responses scoring the minimum): 0 items
- showed poor internal consistency (corrected item-total correlation $<.3$) with other items mapped to that domain: 6 items (4 game-based)
- showed redundancy with the Attention Focusing and Attention Shifting scales of the ECBQ: 2 items

The psychometric properties of the items were also reviewed by age-group. No further items were removed from the main EEFAQ on the basis of this review, but it was identified that 100% of the infants under 12 months performed at floor on the game-based item for the Flexibility scale (“The Sorting Game”). Therefore, a version of the EEFAQ was specifically refined for under 12-month-olds, by excluding The Sorting Game from the Flexibility scale.

As shown in Table 2, after item reduction, internal consistency levels were comparable with the ECBQ scales, and above the .60 threshold for adequate internal consistency commonly used in the literature (Putnam et al., 2006). As described in SM2, items were linked to a wide range of everyday activities and behaviors, with minimal repetition of a target scenario. Cronbach's alpha values for the EEFAQ scales therefore hold up particularly well against those reported for ECBQ scales, which describe a more-restricted range of target behaviors and include some overlap in wording between items (see SM3). However, Cronbach alpha values are likely to be inflated when calculated after removing poorly performing items (Nilsen et al., 2017) and the small sample size means that these indicators can only be considered preliminary. In Study 2, we therefore set out to validate the refined EEFAQ in 2 large independent samples.

TABLE 1 Child age and sex, by age-group, in the Study 1 scale refinement sample

Age in months	9- to 12-month-olds	12-to 18-month-olds	18- to 24-month-olds	24- to 30-month-olds	Combined
Mean (SD)	10.11 (.41)	15.38 (1.66)	20.82 (1.87)	25.55 (.91)	18.07 (4.60)
N (boys)	7 (4)	29 (15)	27 (17)	8 (5)	71 (41)

TABLE 2 Floor and ceiling effects, missingness, and internal consistency of each EEFAQ scale after scale refinement

	Inhibitory control	Flexibility	Working memory	Regulation
Cronbach's alpha	.629	.710	.710	.818
Proportion at floor	0%	0%	0%	0%
Proportion at ceiling	0%	1.5%	1.5%	1.4%

3 | STUDY 2: EEFQ PERFORMANCE AND STRUCTURE

The aims of Study 2 were to establish the psychometric properties of the refined 31-item EEFQ (SM1), in terms of floor and ceiling effects, data missingness, the latent structure of EEFQ data, and measurement invariance by age and sex. We also aimed to demonstrate convergent validity of the EEFQ with existing measures of attentional control and to explore convergent and discriminant associations between the EEFQ and broader aspects of temperament. Finally, we explored the effects of age, sex, and maternal education levels on EEFQ data.

3.1 | Method

3.1.1 | Sample 1 Participants

Parents or guardians of children 8–30 months old were recruited online through Qualtrics research panels between May 9, 2019, and October 20, 2019. Based on recommendations for sample sizes required for Exploratory Factor Analysis (Carpenter, 2018), the minimum target sample size was set at 300 but data collection continued until resource limits were reached. After quality checks detailed in SM 4.1.2, 418 participants contributed data sufficient to compute all 4 EEFQ scales (i.e., a minimum of 70% of items were completed for each scale), a further 10 contributed data sufficient to compute 3 EEFQ scales, a further 16 contributed data sufficient to compute 2 EEFQ scales, and a further 42 contributed sufficient data for 1 EEFQ scale. As detailed in SM 4.1.2, maternal education levels were slightly above the population average, and sample ethnicity was broadly representative of the population. Eighty-two percent of respondents were the child's mother, 17% were the father, and fewer than 1% of respondents were the child's grandmother.

3.1.2 | Sample 2 Participants

Parents or guardians of children 9–30 months old were recruited via the Oxford University BabyLab social media pages and volunteer database between December 7, 2018, and March 19, 2020, either specifically to validate the EEFQ or as part of a larger longitudinal study of EFs; the Oxford Early Executive Functions (OEEF) study—see Table 3. After quality checks detailed in SM 4.1.3, 190 contributed data (via online questionnaire) sufficient to compute all 4 EEFQ scales, a further 42 contributed data sufficient to compute 3 EEFQ scales, a further 23 contributed data sufficient to compute 2 EEFQ scales, and a further 62 contributed data for just 1 EEFQ scale. As detailed in SM 4.1.3, maternal education levels were well above the population average, and sample ethnicity was marginally less diverse than the general population. Ninety-seven percent of respondents were the child's mother, and 3% were the father.

3.1.3 | EEFQ

We used the 31-item EEFQ, refined based on the results of Study 1. In this version of the measure, the instructions and exemplars for game-based items were provided as videos. Videos were filmed using volunteer participants to demonstrate how to administer the games and the range of behaviors that might be expected (see SM1) and embedded in the online version of the measure,

**TABLE 3** Recruitment sources and participant ages for Study 2: Samples 1 and 2

	9–12 months	12–24 months	24–30 months	Combined
Sample 1: Qualtrics Panel participants				
Mean age in months (SD)	10.29 (1.27)	17.93 (3.57)	27.03(1.92)	19.98 (5.41)
<i>n</i> (boys)	20 (12)	345 (165)	121 (34)	486 (211)
Sample 2: OEEF participants (all sources)				
Mean age in months (SD)	10.03 (0.50)	17.56 (3.41)	26.89 (1.55)	14.03 (5.93)
<i>n</i> (boys)	194 (91)	89 (53)	34 (21)	317 (165)
By Sample 2 source				
Sample 2 OEEF pilot				
Mean age in months (SD)	10.43 (0.33)	17.94 (3.55)	26.99 (1.90)	19.95 (5.41)
<i>n</i> (boys)	15 (7)	5 (1)		20 (8)
Sample 2 OEEF main study				
Mean age in months (SD)	9.85 (0.24)	15.86 (0.20)	--	10.04 (1.10)
<i>n</i> (boys)	149 (74)	5 (4)	–	154 (78)
Sample 2 Online study only (EEFQ validation)				
Mean age in months (SD)	10.81 (0.67)	17.66 (3.55)	26.89 (1.55)	18.24 (6.11)
<i>n</i> (boys)	30 (10)	79 (48)	34 (21)	143 (79)

OEEF, Oxford Early Executive Functions study.

accompanied by a written transcript. Games were presented at the beginning of each section (scale), to help to frame respondents' interpretation of the related questionnaire items, but were presented as optional, with the opportunity to defer them to the end of the questionnaire if more convenient; see Table S4.2.3 for details of EEFQ internal consistency when games were excluded. Parents completed the online version of the revised EEFQ, with the exception of 2 participants who requested a print copy which was then entered manually into the online form. On the basis of Study 1 findings, the Sorting Game was excluded from the Flexibility scale in a version of the EEFQ specifically refined for under 12-month-olds and, therefore, was not shown to a subset of Sample 2 participants (those enrolled in the OEEF main study, $n = 170$). Scores for each EEFQ scale were calculated by computing the mean of all items in that scale, where a minimum of 70% of items for that scale were complete (adjusting for the fact that the Flexibility scale had 1 less item in the under-12-month-olds' version). Floor and ceiling effect calculations summarized below include all responses, not just where the 70% inclusion criterion was met, in order to capture floor and ceiling effects among respondents who reported that some items were not applicable. Internal consistency of the individual scales are presented in Table 4.

3.1.4 | IBQ-R VSF and modified attentional control scales from the ECBQ

The most well-established questionnaire measures of attentional control for toddlers are the Attention Focusing and Attention Shifting scales of the ECBQ (Putnam et al., 2006). Attention Focusing items relate to sustained orienting to an object of attention, while Attention Shifting items capture the ability to transfer attentional focus from one activity or task to another. Conventionally for infants 12 months and under the IBQ-R is used, which does not include a

TABLE 4 Cronbach's alpha values, by Measure and Sample

Sample	EEFQ scales				ECBQ scales	
	Inhibitory control	Flexibility	Working memory	Regulation	Attention focusing	Attention shifting
1	.698 <i>n</i> = 409	.540 <i>n</i> = 358	.659 <i>n</i> = 281	.862 <i>n</i> = 406	.594 <i>n</i> = 486	.727 <i>n</i> = 368
2	.717 <i>n</i> = 172	.713 ^a <i>n</i> = 152	.711 <i>n</i> = 129	.821 <i>n</i> = 224	.790 <i>n</i> = 158	.703 <i>n</i> = 148

^aExcluding the Sorting Game as this was not presented to a large subset of Sample 2. Cronbach's alpha for the subset of Sample 2 (*n* = 34) who completed the game = .763.

separate measure of Attentional Shifting. However, as we were interested in understanding how the EEFQ relates to both aspects of attentional control, and wanted to retain consistency of measures across age bands, with kind permission of Samuel Putnam we made minor adaptations to the wording of the ECBQ Attention Focusing and Attentional Shifting scales so they were suitable for 9- to 16-month-olds (see SM3). The original item wording was used with infants aged 16 months and above. Scores for each scale were calculated by computing the mean of all items in that scale, where a minimum of 70% of items for that scale were complete. Internal consistency of the attentional control scales is presented in Table 4.

In addition to the attentional control scales described above, broad measures of temperament were available for a subset of Sample 2 (specifically those infants enrolled in the OEEF study) using the IBQ-R VSF. Scores for the 3 broad temperament factors of Orienting/Regulation, Surgency and Negative Affect were calculated computing the mean of all items for that factor, where a minimum of 70% items was available. Cronbach's alpha for these factors was .643, .620, and .763, respectively.

3.1.5 | Analytic approach

As this is the first investigation into the structure of the EEFQ, we took a two-stage approach. We first explored the factor structure of our Sample 1 data using Exploratory Factor Analysis (EFA) of the summary scores of our a priori scales. We then conducted Confirmatory Factor Analysis (CFA) using our Sample 2 data to confirm the higher order factor structure identified through EFA and to test whether EEFQ data are best described with a second-order model with subscales loading onto a higher order factor or factors, or a first-order model with items loading directly onto a latent factor or factors.

EFA was conducted in RStudio v1.2.5033 using the psych (Revelle, 2019), corpcp (Schafer et al., 2017), and GPA rotation (Bernaards & Jennrich, 2005) packages. Prior to running the EFA we confirmed that the data were suitable for EFA: Bartlett's test of sphericity was significant ($\chi^2(6) = 289.917, p < .001$), and the determinant of the correlation matrix was 0.514. All variables approximated a normal distribution. A Maximum Likelihood factor analysis was then conducted on the 4 scales, using oblique rotation (oblimin).

CFA was conducted in RStudio v1.2.5033 using the lavaan package v0.6–7 (Rosseel, 2012). Although item data were collected on an ordinal likert scale, the number of levels (7) meant that it was appropriate to treat items as continuous data (Rhemtulla et al., 2012). The ML estimator was used to deal with missing data. As shared measurement error is an issue for questionnaire measures (i.e., a single respondent is likely to demonstrate a consistent bias across questionnaire

items which may inflate factor loadings) error variances for all questionnaire items were allowed to correlate. By taking this approach, we can be more confident that the latent factor(s) are driven by EF, not by shared measurement error. Initial model fit was evaluated according to whether values met conventional cutoff values indicating adequate fit: SRMR values close to .08 or below, RMSEA values close to 0.06 (Hu & Bentler, 1999), and CFI values above .90 (Bentler, 1990). Nested model fit indices were compared using a CFI difference test, whereby a difference between CFI greater than or equal to $-.01$ indicates a significant difference in model fit. This approach was chosen in preference to the chi-squared difference of fit test, which is biased by sample size (Cheung & Rensvold, 2002).

To test whether the EEFAQ shows measurement invariance in terms of infant age and sex, we combined Sample 1 and Sample 2 data. We first tested configural invariance (measurement structure is equivalent for each group), followed by weak factorial invariance (factor loadings are equivalent for each group; also known as metric invariance), followed by strong factorial invariance (factor loadings and intercepts are equivalent for each group; also known as scalar or intercept invariance). Again, nested model fit indices were compared using a CFI difference test, whereby a difference between CFI greater than or equal to $-.01$ indicates that measurement invariance has not been achieved. If measurement invariance was not achieved, parameters were gradually released until the conditions of partial measurement invariance could be considered to be met. Subsequent analyses then took a multistage approach to avoid any misspecifications permeating across different parts of the model (McNeish & Wolf, 2020): Factor scores were computed using the preferred (based on model fit indices) model, exported using the `lavPredict` function with default values and then treated as observed data in regression and correlational analyses. To enable future researchers who may be working with smaller samples not suited for CFA to judge whether using composite scores might be expected to yield equivalent results to CFA-derived factor scores, we also repeated these analyses using simple composite scores, as described further in SM4.2.3.

We aimed to establish convergent validity of EEFAQ factor scores by testing for evidence of positive correlations between EEFAQ scores and the Attention Focusing and Attention Shifting scales of the ECBQ. As the strength of associations may be expected to change with development, and as the attentional control scales of the ECBQ have not been previously validated for use with children under 1 year, this analysis was broken down by age band (across both samples). We also aimed to establish that the EEFAQ shows discriminant associations in relation to broader aspects of temperament. Within the subset of Sample 2 10-month-olds with IBQ-R VSF data, we calculated correlations between EEFAQ factor scores and the three broad temperament factors of Orienting/Regulation, Surgency and Negative Affect. Correlations were tested using Pearson correlations with bootstrapping across 1000 samples to estimate confidence intervals.

Finally, we considered the effects of age and sex on EEFAQ data. As these were exploratory analyses, we ran them separately for Samples 1 and 2 to allow us to see whether significant effects replicated across samples. Age effects were tested using linear regression. Sex effects were tested using ANOVA with sex as a fixed factor.

3.2 | Results and discussion

3.2.1 | Data missingness

To evaluate whether the EEFAQ is age-appropriate, we computed the number of cases where more than 2 items (the threshold at which a scale score would not be computed) were marked as not

applicable, for each scale. The scale with the highest proportion of “not applicable” responses was Flexibility (6.5% of 9- to 12-month-olds; 0% for all other age-groups) followed by Inhibitory Control (0.4% of 9- to 12-month-olds; 0% for all other age-groups); see SM 4.2.1 for item details. We next computed the number of cases where infants performed at floor or ceiling on each scale. Floor and ceiling effects were minimal (<1%) for each scale, for each age-group.

3.2.2 | Data structure of the EEFQ

EFA of Sample 1 data showed that 3- and 2-factor EEFQ data models had eigenvalues lower than 0.7; these were rejected based on Jolliffe's criterion (1986), and inspection of the scree plot. A 1-factor model had an eigenvalue of 1.43 and explained 36% of the variance. As shown in Table 5, EFA indicated that the Inhibitory Control, Flexibility and Working Memory scales loaded well above the minimum threshold of 0.4 recommended by Stevens (2002) onto a single factor, which we labeled “Cognitive Executive Function (CEF).” In contrast, the Regulation scale did not load well onto this factor.

Our main aim for the CFA analysis was to investigate the nature of the CEF construct identified by the EFA in a new sample (Study 2, Sample 2). For this first stage, therefore, we included only items mapped to the Inhibitory Control, Working Memory and Flexibility scales. The main question we asked was as follows: Are the cognitive executive functions targeted by the Inhibitory Control, Working Memory and Flexibility scales distinguishable, or do they essentially tap the same underlying construct? Following Miyake et al., (2000), the logic of the analysis was as follows: If items assigned to the Inhibitory Control, Working Memory and Flexibility scales essentially capture the same construct (CEF) then a one-factor model (Model 1) should provide a fit to the data that is statistically no worse than the more complex models described below. If however, Inhibitory Control, Working Memory, and Flexibility are distinguishable constructs, then improved statistical fit (in comparison to Model 1) should be provided by a model in which items load separately onto Inhibitory Control (IC), Working Memory (WM) and Flexibility (FX) factors, which in turn load onto a second-order CEF factor, either freely (Model 2a), or with fixed loadings (Model 2b).

As shown in Table 6, both models showed adequate model fit and there was no significant difference in model fit between either Model 1 and Model 2a (diffCFI = $-.002$), or between Model 1 and Model 2b (diffCFI = $-.001$). Therefore, the first-order CEF model was chosen for reasons of parsimony and EEFQ CEF scores were computed for use in subsequent analyses using Model 1. One-sample Kolmogorov–Smirnov tests showed that CEF factor scores were normally distributed for both Sample 1 ($K-S = .028, p = .200$) and Sample 2 ($K-S = .046, p = .200$). Factor loadings for Model 1 are shown in Table 7. Note that although factor loadings for 5 items did not meet

TABLE 5 Sample 1 Exploratory Factor Analysis factor loadings on the “Cognitive Executive Function” factor

Scale	Factor loadings
Inhibitory control	.58
Flexibility	.81
Working Memory	.66
Regulation	.13

TABLE 6 Model fit statistics for Confirmatory Factor Analysis using Sample 2 EEFAQ data (Inhibitory Control, Working Memory, and Flexibility^a items only)

Model: CEF items only	RMSEA [90% CI]	SRMR	CFI
Model 1: Free loading of items onto unitary CEF factor. Df = 19	.008 [.000, .050]	.021	1.00
Model 2a: Free loading of items onto IC, WM, FX factors, free loading of IC, WM, FX factors onto second-order CEF factor, Df = 16	.026 [.000, .061]	.021	0.998
Model 2b: Free loading of items onto IC, WM, FX factors, fixed loading of IC, WM, FX factors onto second-order CEF factor, Df = 18	.016 [.000, .054]	.021	0.999

IC, Inhibitory Control; WM, Working Memory; FX, Flexibility; Df, degrees of freedom.

^aAs the Flexibility game was only presented to a subset of respondents, including this item caused model convergence issues for Models 2a and 2b. Therefore, the Flexibility game was excluded for the purposes of model comparison.

significance thresholds and thus might be considered to be poor indicators of the latent CEF dimension in this sample (which skewed toward the younger age range—see Table 3, Sample 2) those same items had previously performed well in a different sample (Study 1). Therefore, we chose to retain them for the purposes of validating the EEFAQ structure. We consider measurement invariance by age in detail below.

In SM 4.2.2, we highlight the limitations of using CFA with Regulation items in the current data set. As the Regulation scale is nevertheless of interest theoretically, and to enable readers to contextualize how CEF scores perform against Regulation scores, Regulation scale composite scores (computed from the mean of all Regulation items) are included in regression and correlational analyses reported below. These Regulation scores should be considered vulnerable to measurement error.

3.2.3 | Associations with extant measures of attentional control and temperament

As shown in Table 8, a small positive association was observed between CEF factor scores and Attention Focusing across all age bands. A positive association was also observed between CEF factor scores and Attention Shifting; this association was moderate in the youngest age-group, and large in the oldest age-group. Consistent results were found when composite CEF factor scores were used, although the strength of association was larger in all cases (see Table S4.2.5.1). As shown in Table 8, a weak positive association was observed between Regulation and Attention Focusing, but only among the older two age-groups. There were no significant associations between Regulation and Attention Shifting.

In sum, these results indicate that both CEF and Regulation (from 12 months onwards) scores show convergent validity with Attention Focusing but that only CEF scores are convergent with Attention Shifting.

TABLE 7 Factor loadings for Model 1, Sample 2 data

Latent factor	Indicator item ^a	β
CEF	IC1	0.608***
	IC2	0.706***
	IC3R	0.545***
	IC4	0.341**
	IC5	0.121
	IC6	0.222*
	IC7	0.387***
	IC game	0.503***
	FX1	0.293*
	FX2R	0.074
	FX3	0.652***
	FX4R	0.233*
	FX5	0.547***
	FX6	0.685***
	FX7	0.299**
	FX game	0.751***
	WM1	0.740***
	WM2R	0.359**
	WM3	0.195
	WM4	0.640***
WM5	0.538***	
WM6	-0.005	
WM game	0.135	

***<.001, **<.01, *<.05.

^aSee SM1 for item details corresponding to each code.

As shown in Table 9, in Sample 2 CEF factor scores showed a moderate positive association with Orienting/Regulation and a weak positive correlation with Surgency. In contrast, CEF showed no significant association with Negative Affect. The association between CEF and Orienting/Regulation remained significant after controlling for Surgency ($r(144) = .332, p < .001$ CI: .182 to .468]), but the association between CEF and Surgency was no longer significant after controlling for Orienting/Regulation ($r(144) = .126, p = .130$ CI: -.039 to .278]).

Broadly consistent results were found when composite scores were used; see SM 4.2.5.

A moderate negative association was observed between the Regulation scale and the Negative Affect; see Table 9. No other significant associations with temperament were observed.

In summary, the results of Study 2 indicate that the EEFAQ is appropriate for 9- to 30-month-olds: CEF and separate Regulation scores show minimal floor and ceiling effects, good internal consistency (see also Table S4.2.3), and convergent and discriminant associations with extant measures of attentional control and broader temperament factors. These findings are discussed in more detail in the General Discussion section below.

TABLE 8 Bivariate associations between EEFAQ CEF factor and Regulation scores with ECBQ attentional control scales, by age

	Under 12-month-olds (n = 161–186)	12-to 24-month-olds (n = 326–331)	24- to 30-month-olds (n = 113)
CEF factor scores			
Attention focusing	.234** [.090,.362]	.258*** [.138,.368]	.269** [.060,.451]
Attention shifting	.431*** [.296,.542]	.526*** [.435,.610]	.598*** [.450,.716]
Regulation scale scores			
Attention focusing	-.040 [-.156,.078]	.319*** [.208,.420]	.314** [.091,.526]
Attention shifting	-.007 [-.191,.170]	-.082 [-.204,.045]	.040 [-.180,.243]

Values in square brackets indicate confidence intervals computed using bootstrapping across 1000 samples.

*** $p < .001$, ** $p < .01$, * $p < .05$.

TABLE 9 Bivariate associations between EEFAQ CEF factor and Regulation scores with IBQ-R VSF factor scores (Sample 2 only)

	Orienting/Regulation	Surgency	Negative affect
EEFAQ CEF <i>n</i> = 147	.394*** [.236,.521]	.255** [.108,.402]	.022 [-.146,.190]
EEFAQ Regulation <i>n</i> = 140	.104 [-.050,.250]	.086 [-.001,.081]	-.388*** [-.515,-.261]
Surgency <i>n</i> = 147	.377*** [.241,.502]	–	.100 [-.105,.290]
Negative affect <i>n</i> = 142	.022 [-.130,.171]	.100 [-.105,.290]	–

*** $p < .001$, ** $p < .01$, * $p < .05$.

Values in square brackets indicate confidence intervals computed using bootstrapping across 1000 samples.

3.2.4 | Age effects

As detailed in SM 4.2.4, the EEFAQ CEF factor shows partial measurement invariance for age in our sample. To enable us to identify temporal change in CEF scores beyond those that may be attributable to changes in the structure of CEF over time, age effects were analyzed using factor scores computed to reflect this partial measurement invariance.

As shown in Table 10 and Figure 1, there was a positive association between linear age and CEF factor scores for both Sample 1 and 2. Consistent results were found for composite CEF scores; see SM 4.2.6. Additionally, there was a negative association between the quadratic age term for Sample 2; such that scores rose most-rapidly at the younger-end of the sample age range, followed by a gradual leveling-off of scores (note that the regression coefficients should not be over-interpreted due to multicollinearity of the linear and quadratic age terms (Kraha et al., 2012)—instead, we refer the reader to Figure 1 for interpretation of the age effects). This

TABLE 10 Standardized beta coefficients for regression models predicting CEF factor scores and Regulation scale scores from age and maternal education

Dependent variable	Sample 1			Sample 2		
	Model A	Model B	Model C	Model A	Model B	Model C
CEF factor score						
Age (linear)	0.395***	0.895**	1.159***	0.829***	2.173***	1.660***
Age (quadratic)		-0.504	-0.748*		-1.354***	-0.883**
Maternal education			0.044			-0.198
Adjusted R^2	.156	.157	.177	.688	.718	.639
Regulation						
Age (linear)	0.031	-0.128	-0.374	-0.336***	-1.302**	-1.018*
Age (quadratic)		.098	0.333		0.974*	0.729
Maternal education			-0.018			0.207**
Adjusted R^2			.004			.137

*** $p < .001$, ** $p < .01$, * $p < .05$.

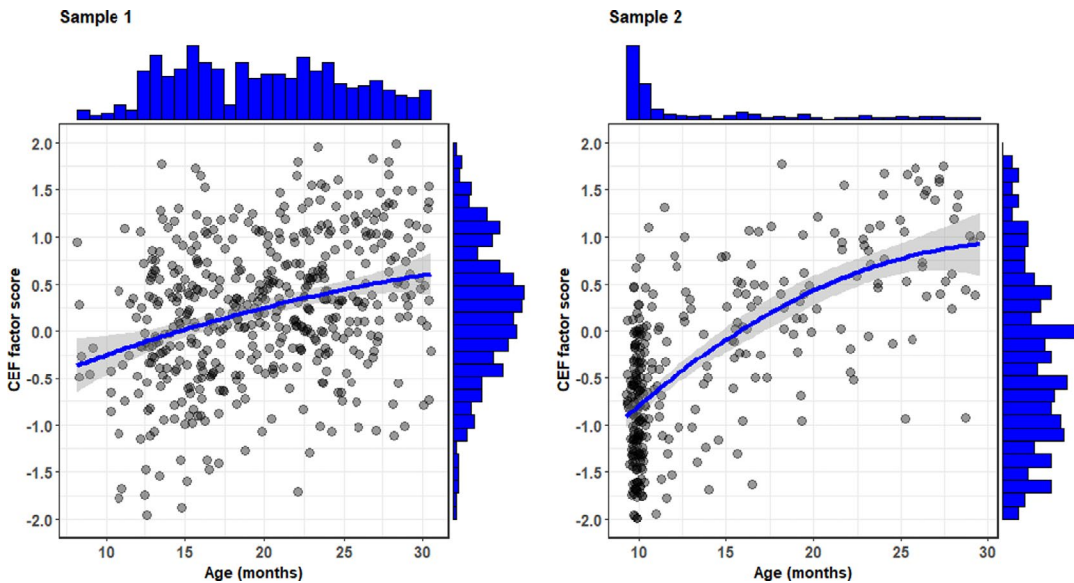


FIGURE 1 CEF factor scores by age, for Sample 1 and Sample 2. Line of best fit shows the quadratic effect of age

quadratic association was also significant for Sample 1, but only when maternal education was included in the model. The quadratic term did not show a significant predictive effect when CEF composite scores were used (see SM 4.2.5).

As shown in Table 10 and Figure 2, there was a negative association between linear age and Regulation scale scores, which was significant for Sample 2 only. Additionally, there was a positive association between the quadratic age term for Sample 2, whereby scores initially decreased in infancy before leveling out in toddlerhood. This quadratic association was no longer significant when maternal education was also included in the model.

3.2.5 | Sex effects

As detailed in SM 4.2.4, the EEFAQ CEF factor shows strong factorial invariance in terms of sex in our sample. As shown in Table 11, although for Sample 1 girls showed higher CEF factor scores ($M = .30, SD = .83$) compared with boys ($M = 0.14, SD = .69$), sex differences accounted for only a very small amount of variance in CEF scores. Further, for Sample 2 there was no significant difference between CEF factor scores for girls ($M = -0.33, SD = .94$) compared with boys ($M = -0.38, SD = .95$), and CEF sex differences did not survive correction for age and maternal education in either sample. When CEF composite scores were used, for Sample 1 (only) girls

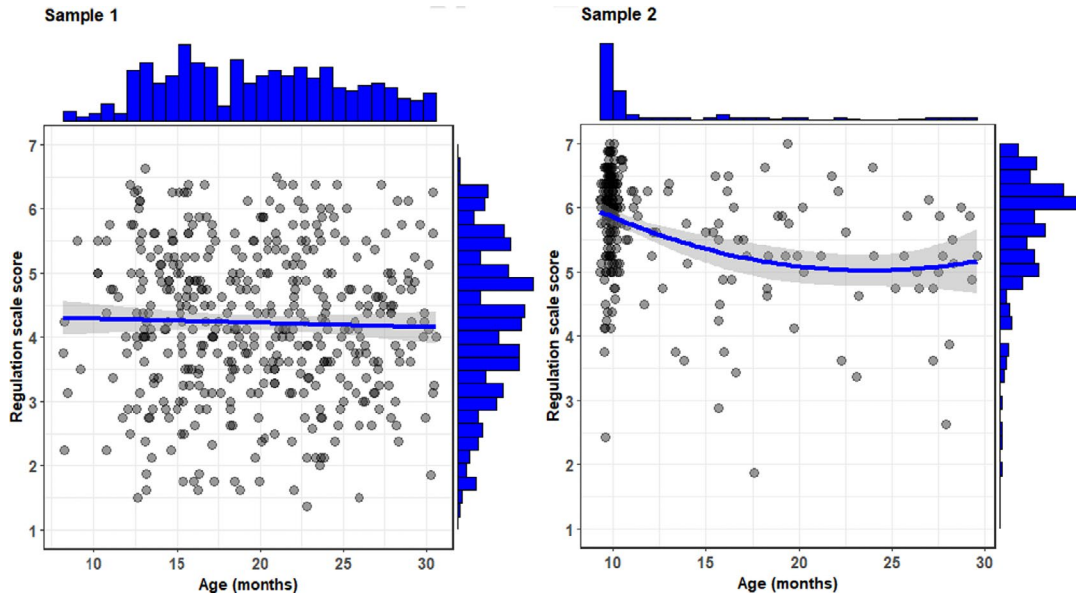


FIGURE 2 Regulation scale scores by age, for Sample 1 and Sample 2. Line of best fit shows the (non-significant) linear effect for Sample 1 and the quadratic effect of age for Sample 2

TABLE 11 Univariate tests of sex differences in CEF factor scores and Regulation scale scores

	Sample 1		Sample 2	
	<i>F</i>	η_p^2	<i>F</i>	η_p^2
CEF factor scores				
Sex	5.517*	.011	0.295	.001
Sex controlling for age (linear and quadratic)	1.766	.004	4.013*	.013
Sex controlling for age and maternal education	2.507	.007	1.081	.006
Regulation scale scores				
Sex	0.790	.002	0.000	.000
Sex controlling for age (linear and quadratic)	0.978	.002	0.101	.000
Sex controlling for age and maternal education	0.024	.000	.000	.000

η_p^2 : partial eta squared.

* $p < .05$.

showed significantly higher scores than boys, even after controlling for age and maternal education; see SM 4.2.7.

There were no significant sex differences in Regulation scores for either Sample 1 (Boys $M = 4.17$, $SD = 1.12$, Girls $M = 4.27$, $SD = 1.21$) or Sample 2 (Boys $M = 5.64$, $SD = 0.95$, Girls $M = 5.64$, $SD = 0.85$); see Table 11.

4 | STUDY 3: SHORT-TERM STABILITY AND LONGITUDINAL AGE-RELATED CHANGE

In this final study, we aimed to establish whether EEFQ scores show homotypic (i.e., within-construct) stability over time and to verify the measure's sensitivity to age-related change, by following up a subset of infants from Study 2. True score change can only be said to occur in the context of longitudinal measurement invariance (Brown, 2015). We therefore first tested for evidence that measurement of the construct of EF using the EEFQ does not change over time within this longitudinal sample. We then investigated the stability of scores and age-related change in three sub-samples which, in combination, encompassed both a narrow and a broad age range, and a test-retest period of 3, 5 or 6 months. This enabled us to propose some insights pertaining to developmental change in very early EFs for follow-up in future research.

4.1 | Method

4.1.1 | Sample 1 Participants

Follow-up data were collected at either 3 (Sample 1a) or 5 months (Sample 1b) after the initial data collection from 124 Sample 1 participants; see Table 12. Respondents were identical between observations in all but 2 cases; in each of these cases, the father contributed the Observation 2 data instead of the mother. These data were included in the analysis below, to provide a conservative estimate of stability.

4.1.2 | Sample 2 Participants

Follow-up data were collected 6 months after initial data collection from 70 Sample 2 participants; see Table 12. Respondents were identical between observations in all but 3 cases; in each

TABLE 12 Age and elapsed time at follow-up (Study 3)

	Sample 1a	Sample 1b	Sample 2
Mean age in months at Observation 1 (SD) [Min, Max]	18.87 (5.24) [10.29, 30.58]	20.47 (4.74) [12.62, 30.15]	9.87 (0.25) [9.30, 10.39]
Mean age in months at Observation 2 (SD) [Min, Max]	22.05 (5.24) [13.31, 33.90]	25.09 (4.75) [17.33, 35.11]	15.88 (0.28) [15.29, 16.67]
Mean elapsed days between Observation 1 and Observation 2 (SD) [Min, Max]	96.67 (7.29) [81, 115]	140.51 (5.31) [130, 156]	181.21 (14.78) [97, 222]
<i>n</i> (boys)	40 (16)	81 (29)	70 (36)

of these cases the father contributed the Observation 2 data instead of the mother. These data were included in the analysis below, to provide a conservative estimate of stability.

4.1.3 | EEFAQ

The 31-item EEFAQ was used for all subsamples. The primary focus of analysis was the CEF factor score computed using the unitary model identified in Study 2, whereby all Inhibitory Control, Flexibility and Working Memory items loaded onto a single factor (or, for the purposes of comparison, a composite score computed by averaging the scores of all Inhibitory Control, Flexibility, and Working Memory items). Additionally, to understand how developmental change in CEF differs from the construct of Regulation (notwithstanding the modeling issues described in Study 2) we computed a Regulation scale score by averaging the score for all Regulation items (where a minimum of 6 items – 70% of all possible items – were complete). Internal consistency of the CEF composite and Regulation scale are presented in Table S4.3.1.

4.1.4 | Analytic approach

CFA was conducted using the analytic approach described for Study 2 with the modification that for this longitudinal dataset residual variances for factor loadings were additionally allowed to correlate over time, to reflect repeated measurement. The computed factor scores for the preferred model were then exported for use in the analyses below. Stability between Observations 1 and 2 was assessed with Pearson correlations, using bootstrapping across 1000 samples to estimate confidence intervals. Because Samples 1a, 1b, and 2 differed significantly in age and in duration between waves, analyses were conducted separately for each sample. To evaluate age-related changes, we ran a repeated measures ANOVA, testing for the main effect of observation point, in the combined sample. As age at Observation 1 varied systematically with the duration of elapsed time between observations, we included time between observations as a covariate. As age at Observation 1 also varied systematically with maternal education, we included maternal education as a covariate. In SM 4.3, we present equivalent analyses using composite scores.

4.2 | Results and Discussion

As detailed in SM 4.3.2, the EEFAQ CEF factor shows partial strong longitudinal measurement invariance (when loadings were freed for 5 items). To enable us to identify temporal change in CEF scores beyond those that may be attributable to changes in the structure of CEF over time, age effects were analyzed using factor scores computed to reflect this partial measurement invariance. In SM 4.3.4, we report the results of equivalent analyses when composite scores were used, and we summarise any differences in conclusions in the text below.

4.2.1 | Stability of CEF and Regulation scores

As shown in Table 13, homotypic associations for the CEF factor were strong for all samples; as shown in Table S4.3.4, consistent results were found when CEF composite scores were

TABLE 13 Bivariate correlations between EEFAQ scores over time, by Study 3 subsample

		CEF factor score Observation 2
CEF factor score Observation 1	Sample 1a	.852 [.760,.927]***
	Sample 1b	.922 [.889,.948]***
	Sample 2	.849 [.775,.905]***
		Regulation scale score Observation 2
Regulation scale score Observation 1	Sample 1a	.454 [.156,.690]**
	Sample 1b	.607 [.474,.738]***
	Sample 2	.413 [.203,.638]***

*** $p < .001$, ** $p < .01$.

used—although the magnitude of association was only moderate for Sample 2. Homotypic associations for the Regulation scale were moderate for all samples (see Table 13).

4.2.2 | Age-related change in CEF scores

In the combined sample, there was a small-to-medium main effect of Observation on CEF factor scores ($F(1,189) = 11.531, p = .001, \eta_p^2 = .058$), a large main effect of Age at Observation 1 ($F(1,189) = 59.997, p < .001, \eta_p^2 = .241$) and a small-to-medium interaction between Observation and Age ($F(1,189) = 10.363, p = .002, \eta_p^2 = .052$). Figure 3 indicates that CEF scores increased between Observation 1 and 2 specifically for younger infants. There was no significant main effect of Elapsed time between Observations ($F(1,188) = 0.268, p = .605, \eta_p^2 = .001$) and no significant interaction between Elapsed time and Observation ($F(1,188) = 2.154, p = .144, \eta_p^2 = .011$)—however, the interaction between Observation and Age was small and only at trend when Elapsed time was included as a covariate in the model ($F(1,188) = 3.138, p = .078, \eta_p^2 = .016$). There was no significant main effect of Maternal education ($F(1,171) = 0.278, p = .598, \eta_p^2 = .002$), no significant interaction between Maternal education and Observation ($F(1,171) = 0.017, p = .896, \eta_p^2 = .000$), and the interaction between Observation and Age remained significant when Maternal education was included as a covariate ($F(1,171) = 6.988, p = .009, \eta_p^2 = .039$). As shown in SM 4.3.4 broadly consistent results were found when CEF composite scores were used but the interaction between Observation and Age remained significant when Elapsed time was included as a covariate in the model. In sum, these results indicate that CEF scores increase with age and that age-related increases may be most pronounced earlier in development (i.e., around the end of the first year of life).

4.2.3 | Age-related change in Regulation scores

In the combined sample, there was a large main effect of Observation on Regulation scale scores ($F(1,181) = 55.169, p < .001, \eta_p^2 = .234$), a medium-to-large main effect of Age at Observation 1 ($F(1,181) = 28.794, p < .001, \eta_p^2 = .137$) and a medium-to-large interaction between Observation and Age ($F(1,181) = 27.940, p < .001, \eta_p^2 = .134$) such that Regulation scores decreased between

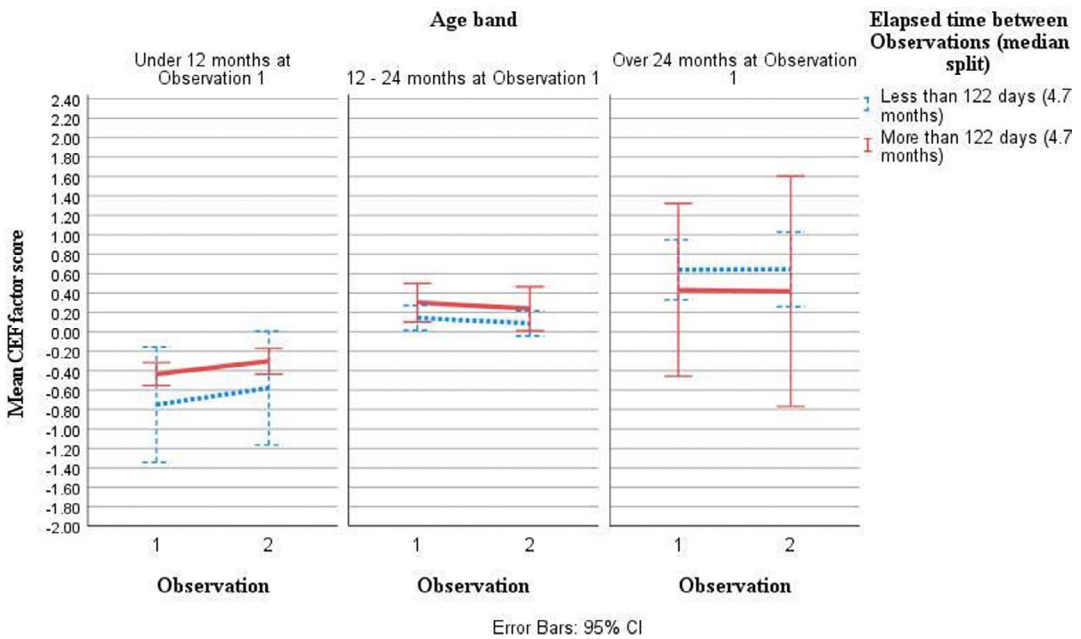


FIGURE 3 Repeated measures CEF factor scores by age band in Study 3

Observation 1 and 2 for the youngest age-group and increased between Observation 1 and 2 for the oldest age-group; see Figure 4. There was a small-to-medium main effect of Elapsed time between observations ($F(1,180) = 5.912, p = .016, \eta_p^2 = .032$) and a small-to-medium interaction between Elapsed time and Observation ($F(1,180) = 6.458, p = .012, \eta_p^2 = .035$), but the interaction between Observation and Age remained significant when Elapsed time was included as a covariate in the model ($F(1,180) = 6.458, p = .012, \eta_p^2 = .035$). There was no significant main effect of Maternal education ($F(1,169) = 0.377, p = .540, \eta_p^2 = .002$), no significant interaction between Maternal education and Observation ($F(1,169) = 2.266, p = .134, \eta_p^2 = .013$), and the interaction between Observation and Age remained significant when Maternal education was included as a covariate ($F(1,169) = 14.089, p < .001, \eta_p^2 = .077$). In sum, these results indicate that Regulation scores decrease around the end of the first year of life and then increase from around the end of the second year of life but that profiles of change are sensitive to the period of time between observations.

5 | GENERAL DISCUSSION

In this series of 3 studies, we have demonstrated that EF-relevant behaviors can be measured through parent report in infants as young as 9 months. In Study 1, we outlined the development and refinement process for a new parent-report measure of EF; the EEFAQ. In Studies 2 and 3 we showed that the EEFAQ affords the opportunity to capture the development of cognitive aspects of EF which are associated with, yet distinct from, commonly measured dimensions of temperament. More cautiously, we suggest that separate regulatory aspects of EF may also be captured by the EEFAQ but that with the current dataset it is difficult to disentangle parents' tendency to rate their child consistently from genuine stability in behavior attributable to an underlying construct.

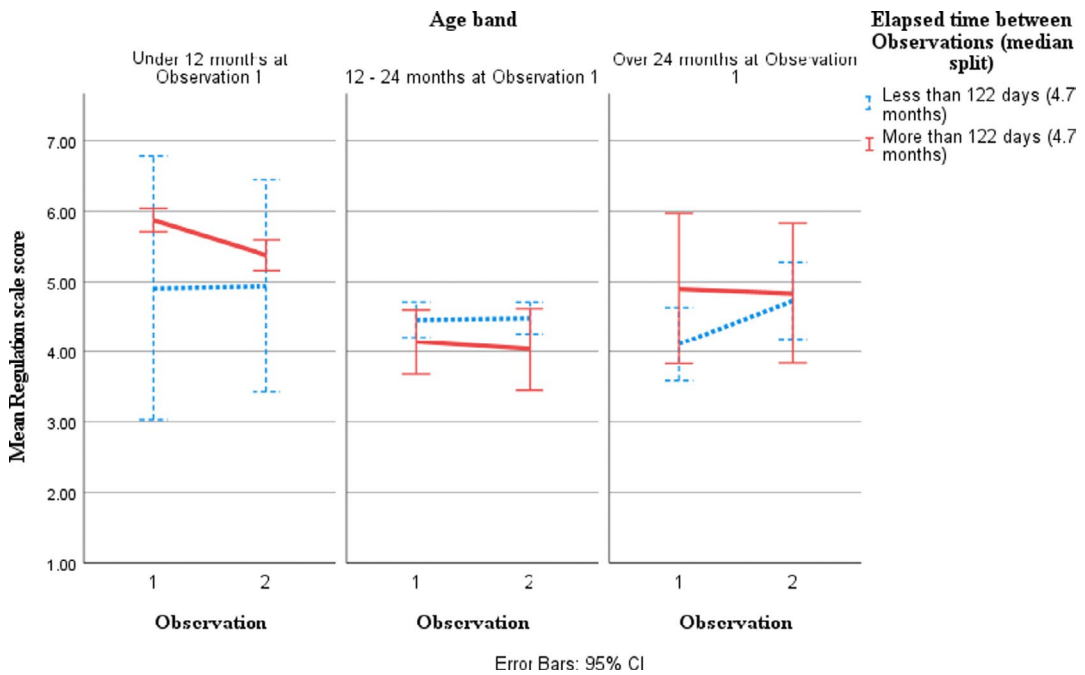


FIGURE 4 Repeated measures Regulation scale scores by age band in Study 3

Unusually for this period of dramatic development, the EEFQ can be used with 9- through to 30-month-olds—enabling researchers to longitudinally assess caregiver-reported EF with a consistent measure which has low resource demands. Below we set out our justification for each of these claims, drawing on the empirical evidence from Studies 2 and 3 and reflecting on broader implications for theory and practice.

In 2 large samples ($n > 300$ for each sample) of UK-based infants, scores from items developed to capture theory-driven aspects of EF (inhibitory control, flexibility, and working memory) load onto a common factor, which we term Cognitive Executive Function (CEF). These findings are consistent with research using performance-based measures of EF with preschool children showing that flexibility skills tend to load with working memory and/or inhibitory skills (Karr et al., 2018). A fourth scale, Regulation, does not fit well with this structure and should be considered separately.

As attentional control is implicated in EF development (Hendry et al., 2016), we expected to find significant positive associations between EEFQ scores and the Attention Focusing and Attention Shifting scales of the ECBQ. Consistent with our hypothesis, both CEF (from 9 months onwards) and Regulation (from 12 months onwards) are positively associated with Attention Focusing. Only CEF scores are convergent with Attention Shifting. Of note, an association was found between CEF and Attention Shifting even in the 9- to 12-month age band, even though Attention Shifting is not conventionally measured at this age, and is not included in the IBQ-R. We recommend therefore that researchers interested in measuring attentional control behaviors in infancy include the adapted Attention Shifting scale when administering the IBQ-R. One interpretation of our pattern of results is that CEF and Attention Shifting (but not Regulation) share common variance relating to cognitive capacity – perhaps in turn relating to processing speed (Hendry et al., 2016). Meanwhile, the association between Attention Focusing and CEF

is consistent with arguments that day-to-day EF-related behaviors entail some engagement of attentional focus. The later onset of the association between EEFQ Regulation and Attention Focusing might indicate that infants do not rely strongly on internal attentional processes to self-regulate until the second year of life.

The dissociation between cognitive (CEF) and regulatory (Regulation) capacity is evocative of the distinction that has been made by some researchers between “cool” and “hot” EF. In previous research, performance measures of both hot (delay of gratification tasks) and cool EF (visual search and working memory tasks) were associated with parent report of inhibitory control and attentional control in toddlerhood (Mulder et al., 2014). An interesting avenue for future research therefore would be to collect EEFQ data alongside a battery of hot and cool lab-based tasks. This would allow one to test the hypothesis that CEF scores correlate with performance on hot as well as cool EF tasks, whereas Regulation scores associate with performance on hot EF tasks alone. Including one or more hot tasks in a CFA model alongside EEFQ Regulation items may also help to resolve the difficulty mentioned above with parceling out common rater bias from the latent construct of Regulation.

In Study 2, we were also able to test convergent and discriminant associations between CEF and Regulation scores and broader dimensions of temperament using the Orienting/Regulation, Surgency and Negative Affect factors of the IBQ-R VSF. Consistent with the associations shown with Attention Focusing (which contributes to the Orienting/Regulation factor), CEF scores showed positive associations with Orienting/Regulation, and also with Surgency. Discriminant associations were found between CEF scores and Negative Affect. Multiple studies have presented evidence for a positive association between aspects of Surgency in infancy (particularly those involving positive affect such as smiling and laughter) and the temperament dimension of Effortful Control in toddlerhood (Casalin et al., 2012; Komsis et al., 2006; Putnam et al., 2008), and Surgency has also been shown to be positively associated with behavioral measures of EF (Frick et al., 2018). In our data, Orienting/Regulation and Surgency showed a moderate positive association. While the association between CEF and Orienting/Regulation remained significant after controlling for Surgency, the association between CEF and Surgency was no longer significant after controlling for Orienting/Regulation. In combination with the literature outlined above, our results indicate that the temperament dimension of Orienting/Regulation is linked to early EF development while Surgency may be implicated in EF development via its association with Orienting/Regulation. As the associations with temperament presented in this study are based on cross-sectional data only, further longitudinal research is required to better understand the direction of causality in these associations.

The EEFQ Regulation scale was, perhaps surprisingly, not significantly associated with the IBQ-R VSF Orienting/Regulation factor. Instead, a modest negative association was observed between the Regulation scale and the Negative Affect factor. In part, this may be attributable to a degree of overlap between the behaviors captured in the EEFQ Regulation scale and the IBQ items relating to Falling Reactivity and Distress to Limitations (which load onto the Negative Affect factor) but is also likely an indication that the EEFQ Regulation scale is sensitive to emotion regulation rather than the more “cognitive” aspects of regulation captured by the Orienting/Regulation factor. Taken as a whole, we conclude from our findings that neither the CEF factor nor the Regulation scale of the EEFQ map neatly onto the Orienting/Regulation factor of the IBQ-R VSF; instead, they capture variation in the cognitive and emotion-regulation aspects of EF and can be used as a complement to existing temperament measures.

We also demonstrate that the EEFQ is appropriate across the 9- to 30-month age range. The CEF factor of the EEFQ and, to a lesser extent (bearing in mind the limitations outlined above)

the Regulation scale, showed good psychometric properties. A model with a unitary CEF factor showed strong factorial measurement invariance for sex and partial strong factorial measurement invariance for age using cross-sectional data. Partial strong longitudinal factorial measurement invariance was achieved for a revised model using 18 of the original 23 CEF items. Further, in both of the cross-sectional datasets in Study 2, and in the longitudinal follow-up data presented in Study 3, Cronbach's alpha values for the CEF composite and Regulation scale were in the range of .751 to .876, indicating good internal consistency. Floor and ceiling effects, and systematic missing data, were low for contributing scales, and for all age-groups, with the exception of the Flexibility scale. Not only is the game-based item included in this scale ("The Sorting Game") too challenging for 9- to 12-month-olds (hence its removal from the refined EEFAQ for under 12-month-olds), 6.5% of the respondents in this age-group marked more than 2 other items of the Flexibility scale as Not Applicable. Nevertheless, the low overall floor effects for Flexibility scale scores, including for the youngest age-group, indicate that even 9-month-olds are able to exhibit some cognitive flexibility. Indeed, as mentioned above, we were able to demonstrate partial strong factorial measurement invariance for age, and Flexibility items were not over-represented compared with items from other scales in terms of not showing strong invariance across age-groups.

CEF factor and composite scores are sensitive to age-related change. In both of the cross-sectional samples described in Study 2, and in the longitudinal sample presented in Study 3, CEF scores showed a significant increase with age. This is consistent with the literature showing age-related improvements in EF (Hendry et al., 2016) and Effortful Control (Putnam et al., 2006) across infancy and toddlerhood. Even after accounting for differences in the demographic profile of the samples (by controlling for maternal education), age explained markedly more variance for Study 2 Sample 2 (which comprised a high proportion of under 1-year-olds) compared with Sample 1 (which comprised a high proportion of 1- to 2-year-olds). Exploratory analysis of the cross-sectional Study 2 data indicated that age showed a quadratic association with CEF scores, whereby scores rose most-rapidly at the younger-end of the sample age range, followed by a gradual leveling-off of scores (Figure 1). Similarly, age-related increases in CEF scores for the Study 3 longitudinal sample were most-evident for participants aged under 12 months at the first observation (Figure 3). Although extant data on trajectories of early EF development are sparse, those data that do exist point toward a similar pattern: the most-rapid age-related improvements in performance on a task tend to occur among the lower end of the age-range covered by that task (Garon et al., 2014; MacNeill et al., 2018; Petersen et al., 2016). This might suggest then that late infancy/early toddlerhood is a time of rapid improvement in EF skills, and by association, a sensitive period for EF development (Taliaz, 2013). Given that targeting interventions within the sensitive period of skill acquisition may potentially yield the greatest opportunity to achieve stable, pervasive change (Luby et al., 2020), the possibility that late infancy/early toddlerhood may be one such sensitive period merits further study in larger, more heterogeneous longitudinal cohorts, using both observer-reported and performance-based measures of EF.

Regulation scores showed an inverse quadratic association with age to that observed for CEF for the cross-sectional Study 2 data (significant only for Sample 2), whereby scores initially decreased in infancy before leveling out in toddlerhood (Figure 2). Similarly, the longitudinal results presented in Study 3 indicated that age-related decreases in Regulation may only be evident in the first year of life—that is, 10-month-old infants tend to show fairly high levels of regulation, which then decline before possibly increasing again toward the end of the second year of life (Figure 4). There is some precedent for this finding in the literature: Age-related increases across the first year of life have previously been reported in the Distress to Limitations scale (a component of Negative Affect, which, as previously discussed shows an inverse association with

Regulation) (Carranza Carnicero et al., 2000; Gartstein & Rothbart, 2003). These changes may be linked to increases in parental limit-setting in response to infants' increased mobility, just as previously reported age-related decreases in Duration of Orienting between 3 and 12 months (Gartstein & Rothbart, 2003) and 6 and 9 months (Carranza Carnicero et al., 2000) may in part be attributable to increased mobility. In later toddlerhood, subsequent increases in Regulation scores—perhaps parallel to increases in Inhibitory Control and Cuddliness scores when the ECBQ is used (Putnam et al., 2006)—might be attributable to a developmental improvement in emotion regulation as well as an increased repertoire of regulatory strategies (Diaz & Eisenberg, 2015). Increases in language ability during this period may also enable children to express their needs and preferences in words, as well as to use language to enhance their self-regulatory strategies (Roben et al., 2013). To better understand the potentially complex interplay of age-related changes in limit setting, language, motor and regulation development, further research combining longitudinal measurement of all these factors is needed.

We found limited evidence of sex differences in EEFAQ scores. For sample 1, girls showed higher CEF scores compared with boys, but sex differences accounted for only a very small amount of variance in CEF scores and after controlling for age and maternal education these differences were only significant for composite and not factor scores. Among sample 2, no significant differences between girls and boys were observed. This inconsistent evidence for sex differences in parent-reported emergent EF echoes the mixed results found in the literature for sex differences in EF among adults—for whom meta-analysis indicates no overall sex difference in any of the domains of performance monitoring, response inhibition, or cognitive set-shifting, but a male advantage in spatial working memory specifically, and a female advantage in delay discounting (Gaillard et al., 2021). The mixed results are also consistent with extant research with toddlers and preschoolers with regards to performance on EF tasks—whereby an advantage is sometimes observed for females, sometimes for males, and sometimes not observed at all (for review, see Hendry et al., 2016)—and with regard to parent-report of EF or effortful control at toddler and preschool age, whereby a significant female advantage has been observed with regard to inhibitory control (Gioia et al., 2002) but not for effortful control, when reported by mothers (Putnam et al., 2006). One possibility is that robust sex differences have not emerged by age 3 years. Another possibility is that performance and observer-report measures suitable for under 3-year-olds do not tend to target those aspects of EF where pronounced sex differences do exist. For example, in the EEFAQ, spatial working memory is represented by only 1 item, and delay discounting is not represented at all; these skills are also not targeted in the BRIEF-P or ECBQ (Gioia et al., 2002; Putnam et al., 2010). However, it should be noted that a female advantage in a wider range of EFs and in effortful control tends to be observed when ratings are provided by teachers (Gioia et al., 2002) or by fathers (Putnam et al., 2006). An interesting avenue for further research therefore would be to consider whether more-consistent sex differences emerge on the EEFAQ when non-maternal ratings are used.

A limitation of the current study is the cultural homogeneity of the samples. In this study, the EEFAQ has been refined with, and validated in, only UK-based samples, with a skew toward high levels of maternal education, particularly for Study 2, Sample 2. It is therefore unclear whether the findings summarized above will generalize to other populations. Although our results indicate that maternal education did not significantly impact age-related changes in parent-reported EF within our samples, future studies, involving larger and more diverse community samples across the target age range are needed to investigate the impact of sociocultural context on the psychometric properties of the EEFAQ.

A further aim for future research was to evaluate how the EEFQ performs among infants with, or at risk for, developmental delay and/or specific EF difficulties—such as infants born pre-term (Mulder et al., 2009) or with Phenylketonuria (DeRoche & Welsh, 2008)—and whether it has predictive validity to later difficulties. Further, although items were designed to minimize language demands, analysis of EEFQ data alongside a broader range of measures is required to establish whether the EEFQ is able to live up to its ambition of distinguishing EF abilities from general cognitive ability among both typically developing and atypically developing infants, including those with language or general cognitive delay.

5.1 | Conclusions

The EEFQ complements measures of infant temperament by providing a carer-report measure more-targeted toward the cognitive capacity and regulatory aspects of EF. The EEFQ is appropriate for use through the ages of 9 to 30 months.

ACKNOWLEDGEMENTS

We would like to thank all the families who participated in data collection for this study. Thanks also to Isobel Greenhalgh and Anna Hall for contributing to data collection, and in particular to Alison Jordan for managing the collection and quality checks of the Sample 2 longitudinal questionnaire data.

We would also like to acknowledge the project funders, as follows: Sample 1 data were supported by a Qualtrics Academic Bursary awarded to A Hendry. Sample 2 data were collected as part of a UK Medical Research Council funded study (MR/N008626/1) led by K Holmboe. A Hendry was supported by an ESRC Postdoctoral Fellowship (ES/S011730/1) and by the Scott Family Junior Research Fellowship in Autism (University College Oxford).

CONFLICTS OF INTEREST

The authors have no known conflicts of interest to disclose.

ORCID

Alexandra Hendry  <https://orcid.org/0000-0003-1985-2521>

Karla Holmboe  <https://orcid.org/0000-0003-3157-6101>

REFERENCES

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71–82. <https://doi.org/10.1076/chin.8.2.71.8724>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65(5), 676–696.
- Bernier, A., Carlson, S. M., Deschênes, M., & Matte-Gagné, C. (2012). Social factors in the development of early executive functioning: a closer look at the caregiving environment. *Developmental Science*, 15(1), 12–24. <https://doi.org/10.1111/j.1467-7687.2011.01093.x>
- Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: early parenting precursors of young children's executive functioning. *Child Development*, 81(1), 326–339. <https://doi.org/10.1111/j.1467-8624.2009.01397.x>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*, 2nd edition. Guilford Publications.

- Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186–197. <https://doi.org/10.1080/15248372.2013.824883>
- Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*(2), 73–92. <https://doi.org/10.1002/icd.298>
- Carpenter, S. (2018). Ten steps in scale development and reporting: a guide for researchers. *Communication Methods and Measures, 12*(1), 25–44. <https://doi.org/10.1080/19312458.2017.1396583>
- Carranza Carnicero, J. A., Pérez-López, J., Salinas, D. C. G., & Martínez-FuentesCarpenter, M. T. (2000). A longitudinal study of temperament in infancy: Stability and convergence of measures. *European Journal of Personality, 14*(1), 21–37.
- Casalin, S., Luyten, P., Vliegen, N., & Meurs, P. (2012). The structure and stability of temperament from infancy to toddlerhood: A one-year prospective study. *Infant Behavior & Development, 35*(1), 94–108. <https://doi.org/10.1016/j.infbeh.2011.08.004>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling, 9*(2), 233–255.
- Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176. <https://doi.org/10.1037/a0019672>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427.
- DeRoche, K., & Welsh, M. (2008). Twenty-five years of research on neurocognitive outcomes in early-treated phenylketonuria: intelligence and executive function. *Developmental Neuropsychology, 33*(4), 474–504. <https://doi.org/10.1080/87565640802101482>
- Devine, R. T., Ribner, A., & Hughes, C. (2019). Measuring and predicting individual differences in executive functions at 14 months: a longitudinal study. *Child Development, 90*(5), <https://doi.org/10.1111/cdev.13217>
- Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development, 56*(4), 868–883. <https://doi.org/10.1111/j.1467-8624.1985.tb00160.x>
- Diaz, A., & Eisenberg, N. (2015). The process of emotion regulation is different from individual differences in emotion regulation: conceptual arguments and a focus on individual differences. *Psychological Inquiry, 26*(1), 37–47. <https://doi.org/10.1080/1047840X.2015.959094>
- Ezpeleta, L., & Granero, R. (2015). Executive functions in preschoolers with ADHD, ODD, and comorbid ADHD-ODD: Evidence from ecological and performance-based measures. *Journal of Neuropsychology, 9*(2), 258–270. <https://doi.org/10.1111/jnp.12049>
- Frick, M. A., Forslund, T., Fransson, M., Johansson, M., Bohlin, G., & Brocki, K. C. (2018). The role of sustained attention, maternal sensitivity, and infant temperament in the development of early self-regulation. *British Journal of Psychology, 109*(2), 227–298. <https://doi.org/10.1111/bjop.12266>
- Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18*(10), 893–900. <https://doi.org/10.1111/j.1467-9280.2007.01997.x>
- Gaillard, A., Fehring, D. J., & Rossell, S. L. (2021). A systematic review and meta-analysis of behavioural sex differences in executive control. *European Journal of Neuroscience, 53*(2), 519–542. <https://doi.org/10.1111/ejn.14946>
- Garon, N., Piccinin, C., & Smith, I. M. (2016). Does the BRIEF-P predict specific executive function components in preschoolers? *Applied Neuropsychology. Child, 5*(2), 110–118. <https://doi.org/10.1080/21622965.2014.1002923>
- Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: sensitivity to age differences. *Child Neuropsychology, 20*(6), 713–736. <https://doi.org/10.1080/09297049.2013.857650>
- Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior & Development, 26*(1), 64–86. [https://doi.org/10.1016/s0163-6383\(02\)00169-8](https://doi.org/10.1016/s0163-6383(02)00169-8)
- Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24–36 months of age. *Developmental Science, 3*(4), 397–404. <https://doi.org/10.1111/1467-7687.00134>
- Ghassabian, A., Herba, C. M., Roza, S. J., Govaert, P., Schenk, J. J., Jaddoe, V. W., Hofman, A., White, T., Verhulst, F. C., & Tiemeier, H. (2013). Infant brain structures, executive function, and attention deficit/hyperactivity

- problems at preschool age. A prospective study. *Journal of Child Psychology and Psychiatry*, 54(1), 96–104. <https://doi.org/10.1111/j.1469-7610.2012.02590.x>
- Gioia, G. A., Espy, K. A., & Isquith, P. K. (2002). *Behavior Rating Inventory of Executive Function – Preschool Version*. Psychological Assessment Resources.
- Hendry, A., Johnson, M. H., & Holmboe, K. (2019). Early development of visual attention: Change, stability, and longitudinal associations. *Annual Review of Developmental Psychology*, 1(1), 251–275. <https://doi.org/10.1146/annurev-devpsych-121318-085114>
- Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, 42, 1–33. <https://doi.org/10.1016/j.dr.2016.06.005>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233–253. <https://doi.org/10.1111/j.2044-835X.1998.tb00921.x>
- Hughes, C., Devine, R. T., Mesman, J., & Blair, C. (2020). Understanding the terrible twos: A longitudinal investigation of the impact of early executive function and parent–child interactions. *Developmental Science*, 23(6), e12979. <https://doi.org/10.1111/desc.12979>
- Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, 35(1), 20–36. <https://doi.org/10.1080/87565640903325691>
- Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, 26(1), 403–422. https://doi.org/10.1207/s15326942dn2601_3
- Jolliffe, I. T. (1986). Principal components in regression analysis. *Principal component analysis* (pp. 129–155). Springer.
- Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, 144(11), 1147. <https://doi.org/10.1037/bul0000160>
- Komsi, N., Raikonen, K., Pesonen, A.-K., Heinonen, K., Keskivaara, P., Jarvenpaa, A.-L., & Strandberg, T. E. (2006). Continuity of temperament from infancy to middle childhood. *Infant Behavior & Development*, 29(4), 494–508. <https://doi.org/10.1016/j.infbeh.2006.05.002>
- Kraha, A., Turner, H., Nimon, K., Zientek, L., & Henson, R. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, 3(44), <https://doi.org/10.3389/fpsyg.2012.00044>
- Luby, J. L., Baram, T. Z., Rogers, C. E., & Barch, D. M. (2020). Neurodevelopmental optimization after early-life adversity: cross-species studies to elucidate sensitive periods and brain mechanisms to inform early intervention. *Trends in Neurosciences*, 43(10), 744–751. <https://doi.org/10.1016/j.tins.2020.08.001>
- MacNeill, L., Ram, N., Bell, M. A., Fox, N. A., & Pérez-Edgar, K. (2018). Trajectories of infants' biobehavioral development: Timing and rate of A-not-B performance gains and EEG maturation. *Child Development*, 89(3), 711–724. <https://doi.org/10.1111/cdev.13022>
- Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, 70(6), 1297–1313. <https://doi.org/10.1111/1467-8624.00095>
- McHarg, G., Ribner, A. D., Devine, R. T., & Hughes, C. (2020). Infant screen exposure links to toddlers' inhibition, but not other EF constructs: A propensity score study. *Infancy*, 25(2), 205–222. <https://doi.org/10.1111/inf.12325>
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52(6), 2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. <https://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>

- Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. M. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology*, *5*, <https://doi.org/10.3389/fpsyg.2014.00733>
- Mulder, H., Pitchford, N. J., Hagger, M. S., & Marlow, N. (2009). Development of executive function and attention in preterm children: A systematic review. *Developmental Neuropsychology*, *34*(4), 393–421. <https://doi.org/10.1080/87565640902964524>
- Nelson, J. M., James, T. D., Choi, H.-J., Clark, C. A. C., Wiebe, S. A., & Espy, K. A. (2016). III. Distinguishing executive control from overlapping foundational cognitive abilities during the preschool period. *Monographs of the Society for Research in Child Development*, *81*(4), 47–68. <https://doi.org/10.1111/mono.12270>
- Nilsen, E. S., Huyder, V., McAuley, T., & Liebermann, D. (2017). Ratings of Everyday Executive Functioning (REEF): A parent-report measure of preschoolers' executive functioning skills. *Psychological Assessment*, *29*(1), 50. <https://doi.org/10.1037/pas0000308>
- Pérez-Edgar, K., Vallorani, A., Buss, K. A., & LoBue, V. (2020). Individual differences in infancy research: Letting the baby stand out from the crowd. *Infancy*, *25*(4), 438–457. <https://doi.org/10.1111/infa.12338>
- Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., & Staples, A. D. (2016). Measuring the development of inhibitory control: The challenge of heterotypic continuity. *Developmental Review*, *40*, 25–71. <https://doi.org/10.1016/j.dr.2016.02.001>
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(35), 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The early childhood behavior questionnaire. *Infant Behavior & Development*, *29*(3), 386–401. <https://doi.org/10.1016/j.infbeh.2006.01.004>
- Putnam, S. P., Jacobs, J. F., Gartstein, M. A., & Rothbart, M. K. (2010). *Development and Assessment of Short and Very Short Forms of the Early Childhood Behavior Questionnaire*. Paper presented at the International Conference on Infant Studies, Baltimore, MD. <Go to WoS>://WOS:000287891800019
- Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development*, *17*(4), 387–405. <https://doi.org/10.1002/icd.582>
- Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research (Version R package version 1.9.12)*. Evanston, Illinois Northwestern University. <https://CRAN.R-project.org/package=psych>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354. <https://doi.org/10.1037/a0029315>
- Roben, C. K. P., Cole, P. M., & Armstrong, L. M. (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Development*, *84*(3), 891–905. <https://doi.org/10.1111/cdev.12027>
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1–36.
- Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. Guilford Press.
- Rothbart, M. K., Ellis, L. K., Rueda, M. R., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality*, *71*(6), 1113–1144. <https://doi.org/10.1111/1467-6494.7106009>
- Rothbart, M. K., & Goldsmith, H. H. (1985). 3 approaches to the study of infant temperament. *Developmental Review*, *5*(3), 237–260. [https://doi.org/10.1016/0273-2297\(85\)90012-7](https://doi.org/10.1016/0273-2297(85)90012-7)
- Rothbart, M. K., & Mauro, J. A. (1990). Questionnaire Approaches to the Study of Infant Temperament. In J. W. Fagen, & J. Colombo (Eds.), *Individual differences in infancy: Reliability, stability, and prediction* (pp. 411–429). Erlbaum.
- Schafer, J., Opgen-Rhein, R., Zuber, V., Ahdesmaki, M., Duarte Silva, A. P., & Strimmer, K. (2017). corpcor: Efficient Estimation of Covariance and (Partial) Correlation.
- Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology*, *26*(1), 445–464. https://doi.org/10.1207/s15326942dn2601_5

- Sherman, E. M. S., & Brooks, B. L. (2010). Behavior Rating Inventory of Executive Function – Preschool Version (BRIEF-P): Test review and clinical guidelines for use. *Child Neuropsychology*, *16*(5), 503–519. <https://doi.org/10.1080/09297041003679344>
- Skogan, A. H., Egeland, J., Zeiner, P., Øvergaard, K. R., Oerbeck, B., Reichborn-Kjennerud, T., & Aase, H. (2015). Factor structure of the Behavior Rating Inventory of Executive Functions (BRIEF-P) at age three years. *Child Neuropsychology*, *22*(4), 472–492. <https://doi.org/10.1080/09297049.2014.992401>
- Skogan, A. H., Zeiner, P., Egeland, J., Urnes, A.-G., Reichborn-Kjennerud, T., & Aase, H. (2015). Parent ratings of executive function in young preschool children with symptoms of attention-deficit/-hyperactivity disorder. *Behavioral and Brain Functions*, *11*, <https://doi.org/10.1186/s12993-015-0060-1>
- Smithson, P. E., Kenworthy, L., Wills, M. C., Jarrett, M., Atmore, K., & Yerys, B. E. (2013). Real world executive control impairments in preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(8), 1967–1975. <https://doi.org/10.1007/s10803-012-1747-x>
- Spiegel, J. A., Lonigan, C. J., & Phillips, B. M. (2017). Factor structure and utility of the Behavior Rating Inventory of Executive Function—Preschool Version. *Psychological Assessment*, *29*(2), 172. <https://doi.org/10.1037/pas0000324>
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*, 4th ed. Erlbaum.
- Taliaz, D. (2013). Skills development in infants: a possible role for widespread neurogenesis? *Frontiers in Behavioral Neuroscience*, *7*(178), <https://doi.org/10.3389/fnbeh.2013.00178>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. <https://doi.org/10.1111/jcpp.12001>
- Vaughn, B. E., Taraldson, B. J., Crichton, L., & Egeland, B. (1981). The assessment of infant temperament: A critique of the Carey Infant Temperament Questionnaire. *Infant Behavior and Development*, *4*, 1–17. [https://doi.org/10.1016/S0163-6383\(81\)80003-3](https://doi.org/10.1016/S0163-6383(81)80003-3)
- Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring development of executive function skills and receptive vocabulary in preschool-aged children: A look at the direction of the developmental pathways. *Infant and Child Development*, *23*(1), 4–21. <https://doi.org/10.1002/icd.1829>
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent Structure. *Developmental Psychology*, *44*(2), 575–587. <https://doi.org/10.1037/0012-1649.44.2.575>
- Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–452. <https://doi.org/10.1016/j.jecp.2010.08.008>
- Willoughby, M. T., & Blair, C. B. (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment*, *28*(3), 319–330. <https://doi.org/10.1037/pas0000152>
- Willoughby, M. T., Wirth, R. J., & Blair, Clancy B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment*, *24*(2), 418–431. <https://doi.org/10.1037/a0025779>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hendry, A., & Holmboe, K. (2021). Development and validation of the Early Executive Functions Questionnaire: A carer-administered measure of Executive Functions suitable for 9- to 30-month-olds. *Infancy*, *00*, 1–30. <https://doi.org/10.1111/infa.12431>