

Longitudinal Stability of Temperament Development from Infancy to Toddlerhood and Implications for Emotional and Behavioral Development: An Empirical Study of a Longitudinal Survey

Jietao LIAN^{*1}, Yi SUN^{*2}, Yuko YATO^{*3}

Abstract:

This study used longitudinal cohort data from a Japanese sample ($N = 110$) to examine developmental trajectories and stability of temperament from infancy (6, 9, and 12 months) to toddlerhood (18 and 24 months), as well as longitudinal associations with socioemotional outcomes assessed by the Strengths and Difficulties Questionnaire (SDQ) at 24 months. Multiple assessments were conducted to evaluate the relative and absolute stability of temperament dimensions. Hierarchical regression analyses, controlling for family background factors, were used to examine associations between temperament and later socioemotional functioning.

Results indicated significant developmental changes around 12 months, with increases in surgency and negative affect followed by declines, and a marked increase in effortful control by 24 months. Patterns of heterogeneous continuity suggested that emerging regulatory capacities may develop alongside early temperamental reactivity. Gender differences were observed, with boys showing higher surgency at 12 months and girls exhibiting higher prosocial behavior at 24 months.

Regression analyses showed that effortful control at 24 months was associated with fewer behavioral regulation difficulties and higher prosocial behavior, whereas negative affect was associated with greater emotional problems. Childcare experience was associated with fewer peer problems. These findings highlight the dynamic nature of early temperament development and its relevance for socioemotional adjustment in early childhood.

Keywords: *Temperament, longitudinal stability, infancy, toddlerhood, behavioral and emotional problems*

*1 Senior Researcher, Research Organization of Open Innovation and Collaboration, Ritsumeikan University

*2 Assistant Professor, Ritsumeikan Asia-Japan Research Organization, Ritsumeikan University

*3 Professor, RARA associate fellow; Dean, College of Comprehensive Psychology; Dean, Graduate School of Human Science, Ritsumeikan University

Email: *1 Lian@fc.ritsumei.ac.jp

Received on 2025/4/30, accepted after peer reviews on 2026/3/5.

1. Introduction

(1) Infant Temperament

Temperament is typically defined as a stable individual trait related to emotional experience and expression, encompassing sensitivity to emotional stimuli, the speed and intensity of emotional reactions, and one's general emotional tone (Rothbart and Bates 2007). These traits are believed to originate from biological foundations and exhibit significant heritability.

Thomas and Chess (1977) offered a broader definition of temperament, emphasizing that it includes not only emotional reactivity but also behavioral traits such as attentional regulation and activity level. They proposed nine fundamental dimensions that constitute the “how” of behavior: activity level, approach–withdrawal, mood, adaptability, threshold of responsiveness, intensity of reaction, distractibility, attention span/persistence, and rhythmicity. They emphasized that the quality of a child's development largely depends on the “goodness of fit” between their temperament traits and environmental demands: high compatibility fosters healthy development, while poor fit may hinder it.

Drawing on existing temperament theories, temperament is regarded as an important individual foundation for early emotional and behavioral development, whereas caregiving and family contexts play a critical role in shaping how these characteristics are expressed and develop over time (Rothbart and Bates 2007; Thomas and Chess 1977). Building on this perspective, Belsky (2016) proposed the differential susceptibility framework, which underscores the dynamic association between individual characteristics and environmental contexts, suggesting that children's developmental trajectories emerge from the ongoing interplay between internal attributes (e.g., temperament, cognitive abilities, and health status) and external conditions (e.g., family and caregiving environments, as well as broader cultural contexts). Empirical studies further indicate that children with different temperament profiles may show distinct developmental patterns across varying family environments (Slagt et al. 2016). In light of these theoretical perspectives, research on temperament development has emphasized the importance of considering caregiving and family background factors when examining associations between temperament and developmental outcomes, either as influential contextual conditions or as variables that should be taken into account analytically.

(2) The Importance of Research on Infant Temperament

A substantial body of research has demonstrated that early childhood temperament is closely linked to subsequent psychological and behavioral development. For instance, a study conducted in France by Abulizi et al. (2017) found that emotional temperament in infancy could predict conduct problems and symptoms of hyperactivity/inattention at 5.5 years of age. Specifically, an active temperament was associated with later externalizing behavior problems, whereas a shy temperament was predictive of future emotional issues. Notably, these associations were highly consistent across individuals and family backgrounds, showing no significant variation based on demographic variables.

Research from China further confirms the influence of temperament on developmental outcomes. Studies have shown a negative correlation between negative affect and psychological resilience in children, whereas effortful control and authoritative parenting were positively associated with resilience. Moreover, negative affect and effortful control were found to indirectly predict resilience through authoritative parenting (Niu et al. 2019). Meanwhile, developmental psychopathology research has identified two temperamental pathways potentially leading to ADHD: one characterized by low effortful control and the other by a strong approach tendency. In contrast, anxiety disorders are more commonly

associated with a combination of high negative affect and low effortful control (Nigg 2006).

Taken together, the existing literature underscores the significant role of temperament in shaping developmental outcomes across cultural contexts. Therefore, early identification of temperament traits in infancy and toddlerhood may provide valuable opportunities for early intervention, promoting more favorable psychological and behavioral development and highlighting the practical importance of research on early temperament.

Further analysis of the literature reveals several key factors influencing the development of temperament in children:

1. Individual Factors

Gender differences are a frequently discussed variable, though findings remain inconsistent. Some studies have reported that boys tend to exhibit higher activity levels, while girls perform better in effortful control (Eiden et al. 2004). In China, girls aged 3 to 5 scored higher in initiative and self-regulation (Niu et al. 2019). However, other studies have found no significant gender differences. For example, research from the United States and South Korea found no gender-based differences in temperament during infancy (Gartstein and Rothbart 2003; Gartstein et al. 2015). These inconsistencies may reflect differences across studies (Cosentino-Rocha et al. 2014; Klein et al. 2013).

2. Age-Related Changes

Studies have shown that temperament characteristics may vary significantly across different age stages in infancy and toddlerhood. For example, Putnam et al. (2008) found systematic differences in temperament dimensions across age groups. Other research suggests that while temperament is relatively stable, it can evolve with maturation and experience. This concept of “dynamic stability” reflects the idea that temperament, while consistent to a degree, can adapt over time and across contexts, demonstrating a continuous pattern of development.

3. Parenting Environment

The early caregiving environment also plays a crucial role in the development of temperament. Maternal mental health, particularly levels of depression and anxiety, is considered a key predictor of early temperament traits. A recent study from the United States found that maternal emotional states at three months postpartum significantly predicted infants’ initial temperament characteristics. After accounting for maternal symptoms, paternal depression and anxiety also showed associations with infant temperament development (Mattera et al. 2023).

Additionally, longitudinal research from Japan has shown that caregiver–infant interaction styles can significantly influence later temperament. For instance, soothing infants through attention-shifting strategies was associated with higher levels of negative affect at 24 months, while infants with greater surgency were more likely to engage in rough-and-tumble play, which was positively correlated with excitability scores at 24 months (Nakagawa and Sukigara 2014).

(3) Limitations of Previous Research and the Purpose of This Study

Previous literature has indicated that boys and girls may differ in certain temperament scores; however, the findings across studies remain inconsistent. Additionally, parents often differ in their evaluations of their children’s temperament. While studies in countries such as the United States, South Korea, the Netherlands, and China have explored the stability of temperament in infancy and early childhood, more empirical research is needed to understand how temperament develops and changes over time within

diverse cultural contexts. For example, the relative stability of longitudinal temperament development can be examined in terms of both homotypic and heterotypic continuity. Homotypic continuity is operationally defined as longitudinal associations of the same temperament dimensions across developmental stages, reflecting the stability of the same constructs over time. In contrast, heterotypic continuity refers to longitudinal associations between different temperament dimensions across developmental stages (Putnam et al. 2008).

In Japan, there is a significant lack of longitudinal empirical data on infant temperament development. Therefore, it is essential to conduct multi-wave longitudinal assessments to further examine the stability of temperament characteristics during infancy and toddlerhood. Although previous research has examined the influence of infant temperament on children's emotional and behavioral development as measured by the Strengths and Difficulties Questionnaire (SDQ) (Abulizi et al. 2017), that study measured temperament only at 12 months. This single time-point measurement is insufficient to evaluate whether temperament at multiple stages in infancy consistently predicts later outcomes from a developmental perspective.

Despite existing longitudinal studies supporting the stability of early childhood temperament, there remains a lack of systematic longitudinal research examining how developmental changes in age and temperament jointly relate to children's early emotional and behavioral development, particularly in non-Western cultural contexts. To address this gap, the present study used longitudinal cohort data from a Japanese sample to examine developmental trajectories and stability of temperament from infancy to toddlerhood, and to investigate longitudinal associations between temperament and children's socio-emotional outcomes, while accounting for individual and caregiving-related factors. Accordingly, the present study addressed the following hypotheses:

Hypothesis 1: Based on previous longitudinal studies, we hypothesized that infant temperament in the Japanese sample would exhibit both relative and absolute stability across infancy and toddlerhood, with different developmental patterns across temperament dimensions.

Hypothesis 2: Given mixed findings in previous research, we explored whether individual factors (e.g., the child's sex) and caregiving environments would be associated with differences in temperament development across time.

Hypothesis 3: Temperament dimensions related to self-regulation and emotional reactivity during toddlerhood were hypothesized to be associated with subsequent emotional and behavioral outcomes measured by the SDQ, after controlling for family background factors.

2. Methodology

(1) Participants

This study is part of the "Ibaraki Cohort Project," a longitudinal study supported by the city of Ibaraki, Osaka Prefecture, which tracks maternal and child development from early pregnancy onwards. The project employs a multi-method approach that includes physiological assessments, behavioral observations, questionnaires, and interviews (Kanzaki et al. 2020).

Data for the present study were collected via questionnaire surveys conducted with a subset of mother-infant pairs participating in the Ibaraki Cohort Project. The survey period spanned from October 2019 to May 2022, targeting mothers with children aged 0 to 24 months. A total of 255 mother-infant pairs were initially recruited.

Regarding temperament assessment, although parental expectations may introduce reporting

bias, previous studies have demonstrated acceptable concordance between questionnaire-based assessments and behavioral observations (Vroman et al. 2014). Therefore, this study adopted maternal self-reports, considering mothers as the primary caregivers, to evaluate infant and toddler temperament characteristics.

In processing the data, responses from the 24-month time point were used as the baseline for constructing the longitudinal sample. A total of 110 cases with valid responses at 24 months were identified and served as the core analytic sample. To construct a longitudinal dataset, questionnaire responses at 6 months ($N = 128$), 9 months ($N = 123$), 12 months ($N = 117$), and 18 months ($N = 116$) were cross-referenced with the 24-month data. Participants who lacked continuous longitudinal follow-up were excluded from the analysis. Among the retained 110 cases, a small proportion of item-level missing values were identified across time points. To address these, multiple imputation was performed to minimize potential bias and maximize the use of available data. As a result, a total of 110 analyzable cases were included in the final longitudinal temperament analysis.

(2) Research Procedure

Between October 2017 and March 2020, the research team distributed informational flyers to residents visiting the Ibaraki City Child Health Center to receive their Maternal and Child Health Handbook, inviting them to participate in the Ibaraki Cohort Project. Recruitment of participants occurred during this period. The questionnaire survey was conducted online. In this study, relevant questionnaires on children's temperament as well as behavioral and emotional development were administered when the children were 6, 9, 12, 18, and 24 months of age, as part of the broader Ibaraki Cohort Project. Children's temperament was assessed using the Japanese translated versions of the Infant Behavior Questionnaire-Revised (IBQ-R) and the Early Childhood Behavior Questionnaire (ECBQ). Behavioral and emotional development problems were measured using the Strengths and Difficulties Questionnaire (SDQ).

As a token of appreciation, participants received a 500-yen Amazon gift card via email. This study protocol was reviewed and approved by the Ethics Review Committee.

(3) Attrition and Follow-up

Attrition is an important consideration in longitudinal research. In the present study, a total of 254 mother-child dyads initially completed study registration. During the follow-up phase, however, some participants did not continue with the questionnaire surveys or participated only in selected components of the study (e.g., behavioral observation). Specifically, 41 dyads did not respond to study-related emails after registration, and an additional 77 dyads discontinued participation after completing the initial baseline survey or explicitly indicated their intention to withdraw (Kanzaki et al. 2022).

Considering the relatively frequent assessments conducted within a limited developmental period, the overall response rate was deemed acceptable in comparison with previous longitudinal studies in Japan reporting comparable questionnaire response rates (e.g., 57.6% at 1 month and 49.7% at 3 months postpartum) and highlighting the challenges of participant recruitment and retention (Murotsu et al. 2020; Ohtani et al. 2006). With regard to representativeness, prior analyses of the Ibaraki Cohort Study have indicated that the number of available consultation partners for parenting is closely associated with continued participation in the survey (Kanzaki et al. 2022).

(4) Measurement

1) Demographic Characteristics of the Household

Theoretical frameworks on temperament suggest that children's temperament may interact with the caregiving environment. Accordingly, in the present study, we collected data on demographic characteristics that might be associated with temperament development. Specifically, children's gender and family environment factors were assessed at 24 months of age.

Children's gender was reported by parents as either "boy" or "girl." After excluding 10 cases with missing data, 40.91% of the sample were identified as boys. In addition, several aspects of the family environment were measured, including whether the family had experienced residential relocation in the past year (yes/no), whether the mother was employed full-time (yes/no), whether there was someone available to consult regarding parenting issues (yes/no), and whether the child was utilizing childcare services (yes/no). In addition, annual household income was reported using an 18-point scale (e.g., 1 = less than ¥700,000, 2 = ¥700,000 to less than ¥1,000,000, ..., 18 = ¥23,000,000 or more). Participants who were unwilling or unable to provide this information could select "I don't want to answer/I don't know."

2) Infant Behavior Questionnaire-Revised (IBQ-R)

To assess infant temperament, we used the Japanese translation of the very short form of the Infant Behavior Questionnaire-Revised (IBQ-R), developed by Putnam et al. (2014). The IBQ-R is based on a three-factor structure and contains 37 items in total: Factor 1: Surgency (13 items); Factor 2: Negative Affect (12 items); Factor 3: Effortful Control (12 items). The Japanese version used in this study was developed with reference to the Japanese short-form translation by Nakagawa and Sukigara (2005).

Items were rated on a 7-point Likert scale ranging from 1 = "never" to 7 = "always," with an additional option "X. No experience, unable to answer" was provided for responses outside the parent's experience. Internal consistency (Cronbach's α) of the IBQ-R for each age point is presented in Table 1. Notably, the Surgency scale at 18 months showed an α value below the commonly accepted threshold of .70.

Table 1. The Cronbach's α of temperament dimensions from infancy to toddlerhood

	6m(N=127)	9m(N=123)	12m(N=117)	18m(N=116)	24m(N=110)
Surgency	.783	.841	.745	.565	.697
Orienting/Effortful	.808	.818	.783	.709	.745
Negative	.748	.812	.777	.742	.806

3) Early Childhood Behavior Questionnaire (ECBQ)

At 24 months, the Early Childhood Behavior Questionnaire (ECBQ) was used to assess toddlers' temperament. The ECBQ has the same three-factor structure as the IBQ-R and demonstrates strong homotypic continuity, which has been supported by prior studies (Putnam et al. 2008). The factors include: Surgency (12 items), Negative Affect (12 items), and Effortful Control (12 items). The Japanese version used in this study was developed with reference to the Japanese short-form translation by Nakagawa et al. (2011).

It also uses the same 7-point response scale as the IBQ-R. In the current study, internal consistency (Cronbach's α) for the ECBQ scales ranged from .697 to .806.

4) Emotional and Behavioral Development

Children's emotional and behavioral development was assessed using the Strengths and Difficulties

Questionnaire (SDQ), a widely adopted and well-validated screening tool designed for children aged 2 to 4 years (Goodman et al. 2000). The Japanese version of the SDQ has demonstrated reliability and practical utility comparable to the original English version (Matsuishi et al. 2008).

In this study, the SDQ was completed by mothers when their child reached 24 months of age. The questionnaire comprises 25 items, which are categorized into five subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. Each item is rated on a 3-point Likert-type scale, with response options of “not true” (0), “somewhat true” (1), and “certainly true” (2), and respondents are instructed to select the option that best describes their child. Each subscale consists of five items, yielding a possible score range of 0–10. These subscales capture a wide range of behavioral and emotional traits, enabling researchers to assess both problem behaviors and strengths in early childhood development.

Internal consistency (Cronbach’s α) for the subscales ranged from .510 to .789. Specifically, the emotional symptoms subscale showed the lowest reliability at $\alpha = .510$, followed by peer problems at $\alpha = .511$, hyperactivity/inattention at $\alpha = .595$, and conduct problems at $\alpha = .579$. In contrast, the prosocial behavior subscale demonstrated acceptable reliability at $\alpha = .789$. When aggregated into broader categories, the internalizing problems composite (emotional symptoms and peer problems) yielded an α of .557, while the externalizing problems composite (conduct problems and hyperactivity) showed a slightly higher reliability at $\alpha = .685$.

Although some subscales showed lower-than-ideal reliability scores, the SDQ remains a valuable tool for screening early behavioral and emotional difficulties in young children, particularly when used alongside other validated developmental measures.

(5) Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics version 25. Because the analytic sample was defined based on the availability of SDQ data at 24 months, no outcome variables were missing. Missing data were observed primarily in temperament measures at earlier time points (all below 6%) and in some background variables, including annual household income (12.7%). The pattern of missing data was considered consistent with the assumption of missing at random (MAR). Missing data were handled using multiple imputation with 20 imputations. All variables included in the main analyses were entered into the imputation model, and pooled estimates were reported for regression analyses following Rubin’s rules (Rubin 1987). Sensitivity analyses based on complete-case data yielded comparable results, indicating that the main findings were robust.

Among demographic variables, child gender was coded as “boy = 1” and “girl = 0,” and the following variables were assessed using a yes/no format: “Relocation in the past year,” “Mother being a full-time worker,” “Availability of someone to discuss parenting issues,” and “Use of childcare services.” Independent-samples t-tests were conducted for each categorical variable to compare the ECBQ and SDQ subscale scores at 24 months between the “yes” and “no” groups.

Furthermore, to examine the consistency (homotypic continuity) between infant temperament (IBQ) and toddler temperament (ECBQ), as well as the relative stability across time points, correlation analyses were performed using pooled data. To assess the absolute stability of temperament, repeated-measures *ANOVAs* were conducted across different age stages for each temperament dimension. Finally, after controlling for demographic characteristics, hierarchical multiple regression analyses were conducted with SDQ subscale scores as dependent variables and temperament at different ages and their changes as independent variables, to examine the longitudinal associations between infant and toddler temperament

on early childhood emotional and behavioral development.

3. Results

(1) Demographic Characteristics of the Household

The results of the analysis of family demographic characteristics are presented in Table 2. To control for the effects of family background variables on outcomes at 24 months of age, we first examined the impact of these variables on toddler temperament (ECBQ) and behavioral development (SDQ) at 24 months. Results from independent-samples t-tests showed that gender differences primarily emerged in the dimension of prosocial behavior, with girls scoring significantly higher than boys, $t = 2.15$, $p < .05$.

Table 2. Demographic characteristics of the household when the child is 2 years old

	<i>N</i>	%		<i>N</i>	%
Gender			Annual household income		
Boy	45	40.91	¥1,500,000 to less than ¥2,500,000	1	0.90
Girl	55	50.00	¥2,500,000 to less than ¥3,500,000	3	2.70
			¥3,500,000 to less than ¥4,500,000	10	9.10
Whether moved within the past year			¥4,500,000 to less than ¥5,500,000	8	7.30
no	91	82.73	¥5,500,000 to less than ¥6,500,000	14	12.70
yes	19	17.27	¥6,500,000 to less than ¥7,500,000	14	12.70
			¥7,500,000 to less than ¥8,500,000	13	11.80
Is there someone I can talk to about childcare issues			¥8,500,000 to less than ¥10,000,000	17	15.50
no	15	13.64	¥10,000,000 to less than ¥12,000,000	7	6.40
yes	95	86.36	¥12,000,000 to less than ¥14,000,000	6	5.50
			¥16,000,000 to less than ¥18,500,000	1	0.90
Whether childcare services are used			¥23,000,000 or more	2	1.80
no	39	35.45	I don't want to answer/I don't know	14	12.70
yes	71	64.55			
			SDQ	<i>M</i>	<i>SD</i>
			Emotional Symptoms	1.35	1.53
Whether the mother is a full-time employee			Hyperactivity/Inattention	2.80	1.56
no	58	52.73	Conduct Problems	2.31	1.72
yes	52	47.27	Peer Problems	2.85	1.80
			Prosocial Behavior	5.33	2.67

Regarding relocation within the past year, children who had experienced a move scored significantly higher on surgency, $t = 2.76$, $p < .05$, and effortful control, $t = 2.72$, $p < .05$, compared to those who had not moved. For the variable on mother's full-time employment status, children of full-time working mothers scored lower on peer relationship problems, indicating fewer issues in peer interactions, $t = -3.17$, $p < .05$.

In contrast, children of mothers who had someone available to consult regarding parenting issues exhibited higher behavioral problem scores, $t = 2.08$, $p < .05$. Additionally, children who had attended

childcare services scored significantly lower on peer relationship problems, $t = -4.15$, $p < .01$, suggesting better peer adaptation.

Furthermore, in the analysis of temperament development during infancy, a gender difference was found only in surgency at 12 months of age, with boys scoring significantly higher than girls, $t = 2.05$, $p < .05$.

(2) Relative Stability/Longitudinal Stability of Temperament

To examine the relative stability of temperament from infancy to toddlerhood, correlation analyses were conducted (see Table 3), involving scores across dimensions of the IBQ and ECBQ.

Table 3. The relative stability of temperament dimensions from infancy to toddlerhood

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.Surgency_6m	1														
2.Effortful_6m	.359**	1													
3.Negative_6m	.405**	.099	1												
4.Surgency_9m	.664**	.367**	.311**	1											
5.Effortful_9m	.222*	.685**	.015	.433**	1										
6.Negative_9m	.247*	-.039	.564**	.430**	.145	1									
7.Surgency_12m	.554**	.246*	.227*	.590**	.258*	.194	1								
8.Effortful_12m	.271**	.670**	.144	.319**	.693**	.103	.313**	1							
9.Negative_12m	.088	-.153	.340**	.183	.001	.599**	.179	-.008	1						
10.Surgency_18m	.545**	.277**	.292**	.506**	.211*	.253**	.535**	.255**	.159	1					
11.Effortful_18m	.199*	.560**	.037	.289**	.568**	.023	.260*	.594**	.039	.490**	1				
12.Negative_18m	.120	-.063	.368**	.084	-.006	.495**	.085	-.142	.641**	.334**	.051	1			
13.Surgency_24m	.383**	.166	.082	.350**	.129	.155	.264**	.113	.010	.384**	.205*	.064	1		
14.Effortful_24m	.227*	.416**	.015	.209*	.451**	.096	.134	.425**	-.006	.316**	.612**	.054	.193*	1	
15.Negative_24m	.196	-.139	.295**	.116	-.077	.079	.254*	.001	.184	.122	-.011	.145	.083	-.145	1

* $p < .05$; ** $p < .01$;

1) Homotypic Continuity

As shown along the diagonals of each subtable in Table 3, significant correlations were observed for surgency and effortful control across all time points, with correlation coefficients ranging from $r = .384$ to $.693$, which were significant and moderate to large in magnitude, indicating moderate to high temporal stability. An exception was found for negative affect, where the correlation between 18 months and 24 months was not statistically significant.

2) Heterotypic Continuity

Positive cross-time associations were also observed between different temperament dimensions. Surgency and effortful control were significantly positively correlated between 6 and 9 months, 9 and 12 months, 12 and 18 months, and 18 and 24 months, with correlation coefficients ranging from $r = .205$ to $.367$. Additionally, surgency and negative affect were significantly positively correlated between 6 and 9 months ($r = .247$, $p < .05$). In contrast, correlations between negative affect and effortful control across different time points were not statistically significant. Within the same time points, significant positive correlations were also found between surgency and effortful control, with correlation coefficients ranging from $r = .193$ to $.490$. However, correlations between negative affect and effortful control within the same periods were not significant. Significant positive correlations between surgency and negative affect were observed at 6, 9, and 18 months, with r values ranging from $.334$ to $.430$.

(3) The Absolute Stability of the Temperament Factors from Infancy to Toddlerhood

Table 4. The absolute stability of temperament dimensions from infancy to toddlerhood

	6m(N=110)		9m(N=110)		12m(N=110)		18m(N=110)		24m(N=110)		P	mean η^2
	M	SE	M	SE	M	SE	M	SE	M	SE		
Surgency	4.369	.070	4.860	.066	5.180	.058	5.002	.061	4.658	.080	**	.268
Orienting/Effortful	5.083	.069	4.981	.071	5.138	.075	4.932	.071	4.938	.076	**	.469
Negative	3.740	.085	4.351	.092	4.759	.093	4.481	.075	3.155	.077	**	.034

** $p < .01$

To examine the absolute stability of temperament development during infancy and toddlerhood, repeated measures *ANOVA* was conducted after multiple imputation ($M = 20$), analyzing changes across three temperament dimensions (surgency, negative affect, and effortful control) at 6, 9, 12, 18, and 24 months of age. The results revealed significant within-subjects main effects for all three dimensions. For surgency, F values ranged from 34.38 to 41.86, $p < .01$, with an average partial eta-squared (η^2) of .268. For effortful control, F values ranged from 86.95 to 97.69, $p < .01$, with an average η^2 of .469. For negative affect, F values ranged from 3.32 to 4.04, $p < .01$, with an average η^2 of .034.

Developmental patterns showed that surgency continuously increased prior to 12 months of age, while effortful control exhibited a gradual decline between 12 and 18 months. Negative affect increased before 12 months but showed a decreasing trend from 12 to 24 months. As illustrated in Figure 1, each of the three temperament traits demonstrated significant developmental shifts around 12 months of age, suggesting that this period may represent an important turning point in early temperament development.

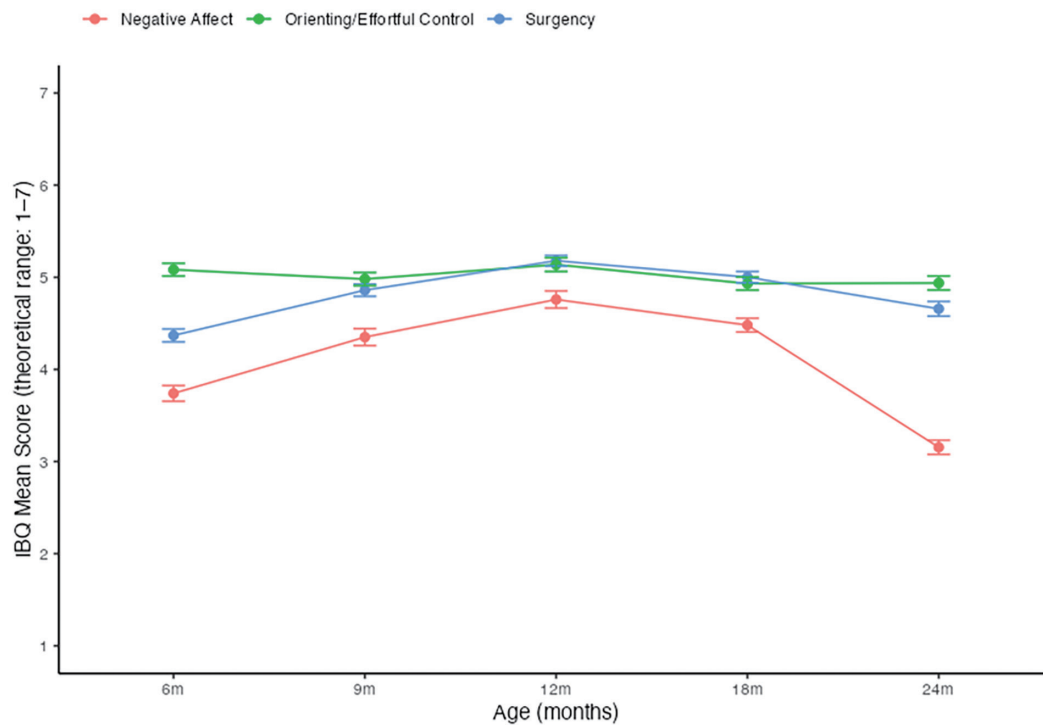


Figure 1. The absolute stability of the temperament factors from infancy to toddlerhood

(4) Longitudinal Associations Between Children's Temperament and SDQ Scores at 24 Months Across Infant Periods

Table 5 Results of Hierarchical Multiple Regression Analysis

	ES			CP			HI			PP			PB		
	B (SE)	95% CI		B (SE)	95% CI		B (SE)	95% CI		B (SE)	95% CI		B (SE)	95% CI	
Gender	-0.34 (0.28)	[-0.88, 0.21]		-0.19 (0.31)	[-0.80, 0.42]		-0.39 (0.30)	[-0.98, 0.21]		0.42 (0.33)	[-0.23, 1.06]		-0.69 (0.51)	[-1.68, 0.31]	
Whether moved within the past year	0.14 (0.37)	[-0.59, 0.87]		-0.24 (0.43)	[-1.08, 0.60]		0.01 (0.40)	[-0.77, 0.80]		-0.16 (0.44)	[-1.02, 0.71]		0.87 (0.66)	[-0.42, 2.16]	
Whether the mother is a full-time employee	-0.05 (0.32)	[-0.68, 0.59]		-0.49 (0.35)	[-1.18, 0.19]		-0.22 (0.34)	[-0.88, 0.44]		-0.26 (0.38)	[-1.00, 0.49]		0.33 (0.55)	[-0.75, 1.40]	
Whether childcare services are used	-0.09 (0.33)	[-0.74, 0.56]		0.24 (0.36)	[-0.46, 0.95]		-0.32 (0.34)	[-0.99, 0.36]		-1.29 (0.39)	[-2.05, -0.53]	**	0.18 (0.56)	[-0.92, 1.29]	
Is there someone I can talk to about childcare issues	0.21 (0.40)	[-0.58, 1.00]		0.87 (0.43)	[0.02, 1.72]	*	0.22 (0.41)	[-0.59, 1.03]		0.06 (0.46)	[-0.85, 0.97]		0.44 (0.68)	[-0.89, 1.76]	
Annual household income	0.02 (0.04)	[-0.06, 0.10]		0.06 (0.05)	[-0.03, 0.15]		-0.04 (0.04)	[-0.13, 0.04]		-0.08 (0.05)	[-0.18, 0.01]		0.05 (0.06)	[-0.07, 0.18]	
Surgency_6m	-			-											
Effortful_6m	-			-											
Negative_6m	-			-											
Surgency_12m	-			-											
Effortful_12m	-			-											
Negative_12m	-			-											
Surgency_24m	-			0.56 (0.18)	[0.20, 0.92]	**	-			-					
Effortful_24m	-			-0.88 (0.20)	[-1.26, -0.49]	**	-0.76 (0.18)	[-1.12, -0.40]	**	-			1.39 (0.30)	[0.79, 1.98]	**
Negative_24m	0.82 (0.17)	[0.48, 1.15]	**	-											
ΔSurgency6-12	-			-											
ΔEffortful6-12	-			-											
ΔNegative6-12	-			-											
ΔSurgency12-24	-			-											
ΔEffortful12-24	-			-											
ΔNegative12-24	-			-											
R ²	.21 – .35			28 – .33			.24 – .35			.19 – .25			.24 – .32		

* $p < .05$; ** $p < .01$; Emotional Symptoms: ES; (Conduct Problems: CP) (Hyperactivity/Inattention: HI) (Peer Problems: PP) (Prosocial Behavior: PB)

Hierarchical multiple regression analyses were conducted to examine the longitudinal associations between temperament at 6, 12, and 24 months of age, as well as changes across these periods, on the different subscales of the SDQ at 24 months. Given the potential influence of family demographic variables observed in previous analyses, these variables were controlled for in the regression models. Specifically, in Step 1, family environment variables (e.g., gender, use of childcare services, availability of parenting support) were entered using forced entry. In Step 2, scores for temperament dimensions at 6, 12, and 24 months, along with the change scores between 6–12 months and 12–24 months, were entered using the stepwise method. The selection of change scores was informed by the developmental trajectories shown in Figure 1, where the most substantial changes occurred around 12 months. Therefore, changes before and after 12 months were included as predictors to test the effects of temperament development on SDQ outcomes. Prior research has suggested that using difference scores as predictors in regression models can often be a preferable approach (Dalecki and Willits 1991).

The regression results are presented in Table 5. After controlling for family variables (Step 2), several predictors showed discrepancies compared to the t-test results. Specifically, gender, relocation within the past year, maternal full-time employment status, and annual household income were not significantly associated with SDQ problem scores or prosocial behavior at 24 months. Consistent with the t-test findings, however, children who had used childcare services showed significantly lower scores on peer problems ($B = -1.29$, $SE = 0.39$, $p < .01$). In addition, having someone to talk to about childcare-related concerns was significantly associated with higher levels of conduct problems ($B = 0.87$, $SE = 0.43$, $p < .05$).

Regarding longitudinal associations between temperament and SDQ outcomes, emotional problems were significantly associated only with negative affect at 24 months ($B = 0.82$, $SE = 0.17$, $p < .01$). Conduct problems at 24 months were significantly associated with both surgency ($B = 0.56$, $SE = 0.18$, $p < .01$) and effortful control ($B = -0.88$, $SE = 0.20$, $p < .01$), such that higher levels of surgency were associated with more conduct problems, whereas higher levels of effortful control were associated with fewer conduct problems. For hyperactivity/inattention, only effortful control at 24 months showed a significant association ($B = -0.76$, $SE = 0.18$, $p < .01$). No significant associations were found between temperament and peer problems at any developmental stage. With regard to prosocial behavior, effortful control at 24 months was significantly and positively associated with prosocial behavior ($B = 1.39$, $SE = 0.30$, $p < .01$).

4. Discussion

(1) Longitudinal Stability of Temperament

This study investigated the developmental stability of temperament from infancy to toddlerhood using the IBQ and ECBQ. Consistent with previous research, the results confirmed the presence of homotypic continuity in temperament development, particularly across the key developmental stages at 6, 9, 12, and 18 months of age. In line with prior findings (Putnam et al. 2006; Sieber and Zmyj 2022), the present study also found that surgency exhibited low stability. Additionally, negative affect showed relatively low stability between the IBQ and ECBQ assessments, which may be attributed to significant changes in this dimension observed at 24 months—a finding further supported by the absolute stability analyses. Previous studies have shown high correlations of temperament traits across different time points, supporting their homotypic continuity (e.g., the consistency of negative affectivity scores over time) (Komsis et al. 2006; Putnam et al. 2008). Additionally, research has indicated that infant surgency

can predict effortful control in toddlerhood, supporting the existence of heterotypic continuity, i.e., the developmental continuity across different dimensions of temperament (Casalin et al. 2012).

The results not only replicated earlier findings but also extended them by demonstrating that temperament traits in a Japanese infant and toddler sample exhibit a certain degree of longitudinal stability. Furthermore, the study provided evidence for the existence of heterotypic continuity. Specifically, significant positive correlations between effortful control and surgency were consistently observed across different age periods. This suggests that during the early stages of prefrontal cortex development, emotional reactivity (surgency) and self-regulatory abilities (effortful control) may develop in tandem. More specifically, infants who are more active and outgoing may have greater opportunities for social interaction and task engagement, which could, in turn, foster the development of regulatory capacities such as attentional control and inhibitory skills (Rothbart and Bates 2007).

These findings highlight the importance of considering heterotypic continuity when understanding the dynamic processes of temperament development.

At the same time, no significant correlations were found between negative affect and effortful control either within the same time points or across different time points. Prior research has shown that correlations between negative affectivity and effortful control tend to be weak or absent in the first year of life but become more robust after 18 months (Putnam et al. 2008). It has also been suggested that such associations stabilize further around ages 3 to 4 (Rothbart and Bates 2007). However, in the present study, no significant correlation between negative affect and effortful control was observed, even between 18 and 24 months. This finding not only supports the idea that negative affect and effortful control are structurally distinct dimensions in early development but also suggests that the maturation of effortful control might not be sufficiently advanced before 24 months of age. Consequently, negative affect and effortful control may represent relatively independent temperament components at least prior to 24 months. For infants and toddlers under 24 months, high levels of negative affect and weak attentional control may require different types of supportive strategies.

Consistent with the findings of Putnam et al. (2008), which indicated that infant surgency can predict later effortful control, the present study confirmed this association at 6, 9, and 18 months. However, no significant relationship was found at 12 months, possibly reflecting a period of significant transformation in the temperament characteristic of surgency around that age.

(2) The Absolute Stability of the Temperament Factors from Infancy to Toddlerhood

This study further explored the absolute stability of temperament from infancy to toddlerhood. The results revealed dynamic changes in temperament scores across different ages, supporting the notion that “temperament stability is dynamic” as proposed by Bornstein et al. (2015). Notably, significant changes in temperament traits were observed around 12 months of age: both surgency and negative affect increased before 12 months and then declined thereafter. This pattern may be associated with physiological developmental processes in infancy (Sieber and Zmyj 2022). The developmental trajectory suggests that infants may experience heightened emotional reactivity before 12 months, followed by a gradual maturation of regulatory functions, leading to increased emotional stability as they grow older.

Additionally, the present study found that negative affect peaked at 12 months of age, a trend consistent with findings from previous Japanese studies (Nakagawa and Sukigara 2014). The increase in negative affect during this period may correspond to critical milestones in social-emotional development. For example, around 6 months of age, infants begin to develop object permanence, leading to emotional reactions to caregiver separations—commonly referred to as “separation anxiety” (Stern 2018). Similarly,

by around 9 months of age, “stranger anxiety” typically emerges, reflecting heightened sensitivity to unfamiliar people and environments, which may further exacerbate the expression of negative emotions (Sroufe 1977). These social challenges likely contribute to the elevated emotional reactivity observed around the 12-month mark, manifesting in greater irritability, fussiness, and emotional dysregulation.

Importantly, the study also observed a significant decline in negative affect scores from 12 to 24 months. This phenomenon may reflect the gradual maturation of the infant’s self-regulatory system during the second year of life. Prior research has indicated that as prefrontal cortex functions and executive control abilities develop, infants begin to acquire early emotion regulation skills such as delay of gratification, attentional shifting, and emotional inhibition (Bernier et al. 2010; Rothbart et al. 2011). The acquisition of these regulatory capacities enables infants to more effectively manage negative emotional responses in the face of frustration or discomfort, resulting in more stable emotional behaviors.

In summary, the development of negative affect during infancy and toddlerhood appears to follow a dynamic trajectory characterized by an initial increase followed by a decline. This developmental pathway may be driven by the interplay between physiological maturation and social contextual factors. Thus, understanding changes in infants’ emotional reactivity requires an integrated perspective that considers both neuropsychological development and critical social experiences.

(3) The Effects of Household Demographic Variables and Temperament on Children’s SDQ

The findings of this study revealed that, consistent with previous research, there are gender differences in children’s emotional and behavioral development. Prior literature has suggested that gender differences emerge gradually during the preschool years; for example, girls were found to demonstrate superior prosocial behavior by 5.5 years of age (Abulizi et al. 2017). Extending these findings, the present study observed that gender differences had already become evident by 24 months, with girls scoring significantly higher than boys on prosocial behavior scales. This suggests that gender-related differences in prosocial abilities may emerge earlier than previously recognized.

However, it is important to note that when family background variables (e.g., family income) and child temperament characteristics were controlled for in the hierarchical regression model, the longitudinal associations between gender and prosocial behavior were no longer significant. This result implies that gender differences may be partially mediated or moderated by individual temperament traits and family socialization factors. In other words, girls’ early advantages in prosocial behavior may not solely reflect biological differences but may also stem from differing developmental trajectories in temperament and parenting practices. For instance, research has shown that parents tend to encourage emotional expression and caregiving behaviors in girls while emphasizing autonomy and control in boys (Leaper 2002). Such socialization experiences may reinforce the observed gender differences in prosocial development.

Regarding gender differences in temperament, the present study found that only at 12 months of age did boys score significantly higher than girls in surgency, consistent with earlier findings (Ruddy and Bornstein 1982). Prior studies have also reported that girls tend to exhibit better orienting/regulation capacities during infancy (Casalin et al. 2012), and greater persistence by 9 months of age (Ogura 2015). These findings suggest that gender differences in temperament development are not uniform across developmental stages but may be moderated by age and developmental timing, indicating a potential interaction between gender and age. Moreover, gender differences are not solely rooted in biological bases but may also be shaped by sociocultural factors. Research has pointed out that parental

gendered expectations, cultural norms, and differential parenting practices can all influence temperament expression and developmental trajectories in infants and toddlers (Casalin et al. 2012; Desmarais et al. 2019; Sieber and Zmyj 2022). Therefore, when examining gender differences in temperament development, it is essential to consider the combined influences of biological maturation, developmental stage, and sociocultural context. Future studies should explore the mechanisms underlying these interactions to better elucidate gender-specific pathways in temperament development.

In addition, this study further examined whether family background and caregiving factors were longitudinally associated with children's socio-emotional development, as measured by the SDQ subscales, using hierarchical regression analyses. Some predictors showed inconsistencies compared to the prior t-test results. For example, variables such as gender, relocation within the past year, maternal full-time employment status, and family income were not significantly associated with SDQ total problem scores or prosocial behavior after controlling for other factors in the regression models.

This discrepancy may indicate that the apparent significance of certain family variables in univariate analyses reflects shared variance with other factors, which diminish when multiple predictors are simultaneously considered. In contrast, some variables demonstrated consistent associations across both t-tests and regression analyses. Specifically, children who had attended childcare services exhibited significantly fewer peer problems, a finding consistent with previous research highlighting the role of childcare experiences in early social development. Childcare settings often provide opportunities for peer interaction, allowing children to practice skills related to rule understanding, conflict resolution, and cooperative behavior, which may be associated with fewer peer-related difficulties (NICHD Early Child Care Research Network 2003).

Furthermore, the regression analyses indicated that having someone to talk to about childcare-related concerns was significantly associated with higher levels of conduct problems, although no significant associations were observed with other SDQ subscales or prosocial behavior. This finding does not necessarily suggest a negative effect of parenting support itself, but may instead reflect differences in maternal perception or reporting of children's behavioral difficulties. Mothers experiencing greater caregiving stress or behavioral management challenges may be more likely to seek opportunities for discussion and support. At the same time, having such outlets may increase mothers' sensitivity to children's behavioral problems, leading to higher reported levels of conduct difficulties.

Regarding the influence of temperament on behavioral development, the study first found that negative affect at 24 months was significantly associated with emotional problems, with higher negative affect predicting greater emotional difficulties. However, no significant associations were found between early (6–12 months) temperament and later emotional symptoms, possibly reflecting the developmental maturation of emotion regulation abilities by 24 months. As noted earlier, negative affect scores declined at 24 months, suggesting a diminished influence of early temperament traits on current emotional outcomes. This finding is consistent with the research by Rothbart et al. (2011), which emphasized that changes in brain connectivity from infancy to toddlerhood can alter the predictive power of early emotional reactivity for later internalizing problems. Thus, the present study highlights the importance of continuously monitoring current emotional reactivity during early development, as persistently high negative affect may indicate an elevated risk of internalizing symptoms.

Regarding conduct problems, both surgency and effortful control at 24 months were found to be significant predictors. This finding aligns with Abulizi et al. (2017), who reported that higher levels of surgency during infancy were associated with conduct problems at 5.5 years. The results suggest that early high activity and impulsivity may constitute long-term risk factors for behavioral difficulties.

Notably, effortful control appeared to play a moderating or buffering role; children with higher surgency levels but stronger attentional control and inhibitory abilities exhibited a lower risk of developing conduct problems. These results emphasize the interactive mechanisms among different temperament dimensions and underscore the protective role of effortful control in behavioral development. They also suggest that fostering self-regulation skills may be particularly beneficial for high-surgency children in preventing behavior problems.

For hyperactivity/inattention, the findings indicate that effortful control at 24 months was uniquely associated with individual differences in attention problems and hyperactive behaviors. No significant predictors were found for peer problems from early temperament measures, indicating that the development of peer relationships likely involves more complex social skills and context-specific interactions that may not be directly predicted by early temperament traits (Denham et al. 2003). Finally, regarding prosocial behavior, the findings suggest that higher effortful control at 24 months significantly predicted greater prosocial tendencies, such as empathy and helping behaviors.

The present findings should be interpreted within the cultural and caregiving context of Japan. Previous research has shown that Japanese parenting practices place strong emphasis on emotional regulation, behavioral restraint, and harmonious social adjustment from early childhood (Rothbaum et al. 2000; Trommsdorff and Friedlmeier 2010). Within such a cultural framework, individual differences in self-regulatory capacity may be particularly salient for children's socio-emotional development.

In this context, the observed associations between effortful control at 24 months and multiple SDQ outcomes—including fewer behavior problems, lower levels of hyperactivity/inattention, and higher prosocial behavior—may reflect the alignment between children's regulatory capacities and culturally valued modes of emotional and behavioral expression. Children who exhibit higher levels of effortful control may be better able to adapt to caregiving environments that prioritize emotional moderation and social coordination, which could be associated with more favorable socio-emotional adjustment in early childhood.

In conclusion, the results of this study highlight the crucial role of infant and toddler temperament traits, particularly effortful control and negative affect, in predicting early emotional and behavioral outcomes as well as social competencies. They also reveal distinct pathways through which different temperament dimensions influence various developmental domains. Future research should integrate observational data and physiological indicators to further investigate the interactive effects of temperament and social environments on socio-emotional development.

(4) Implications for Future Research

Based on data from a cohort study and employing longitudinal analysis methods, the present study provided a more detailed division of developmental stages during infancy and toddlerhood compared to previous research. It examined the developmental trajectories of temperament traits at each stage and their impact on early childhood emotional and behavioral problems. For example, some previous studies have focused solely on the significant associations between temperament at 12 months of age and later emotional and behavioral outcomes (Abulizi et al. 2017). However, the present study found that temperament traits underwent significant changes around 12 months when examining the absolute stability of temperament, suggesting that using temperament characteristics at 12 months alone may not adequately represent the developmental patterns throughout infancy.

Furthermore, the observed variations in temperament traits across different stages provide important theoretical bases for early intervention. Identifying critical periods of temperament

transformation may offer more precise timing for the prevention and intervention of children's psychological and behavioral problems. Given that most existing studies have employed cross-sectional designs, examining the association between infant temperament and socio-emotional development at a single time point, this study adds to the literature by providing longitudinal evidence. This longitudinal approach contributes to a deeper understanding of the developmental processes of early temperament and its implications for subsequent adjustment.

Nevertheless, several limitations should be acknowledged. Although the overall internal consistency of the SDQ total score was acceptable, some subscale scores yielded internal consistency values below 0.60, which may affect the reliability and replicability of certain findings. In addition, future work should consider analytic approaches that explicitly model within-person developmental trajectories. Therefore, further accumulation of longitudinal data is needed to support the construction of more refined path analysis models in future research.

References

- Abulizi, Xian, Laura Pryor, Grégory Michel, Maria Melchior, Judith van der Waerden, and EDEN Mother–Child Cohort Study Group. 2017. Temperament in Infancy and Behavioral and Emotional Problems at Age 5.5: The EDEN Mother-Child Cohort. *PLoS One*, 12(2), e0171971. <<https://doi.org/10.1371/journal.pone.0171971>>
- Belsky, Jay. 2001. Emanuel Miller Lecture: Developmental Risks (Still) Associated with Early Child Care. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(7), 845–859. <<https://doi.org/10.1111/1469-7610.00782>>
- . 2016. The Differential Susceptibility Hypothesis: Sensitivity to the Environment for Better and for Worse. *JAMA Pediatr*, 170(4):321–322. <[doi:10.1001/jamapediatrics.2015.4263](https://doi.org/10.1001/jamapediatrics.2015.4263)>
- Bernier, Annie, Stephanie M. Carlson, and Natasha Whipple. 2010. From External Regulation to Self-Regulation: Early Parenting Precursors of Young Children's Executive Functioning. *Child Development*, 81(1), 326–339.
- Bornstein, Marc H., Diane L. Putnick, Maria A. Gartstein, Chun-Shin Hahn, Nancy Auestad, and Deborah L. O'Connor. 2015. Infant Temperament: Stability by Age, Gender, Birth Order, Term Status, and Socioeconomic Status. *Child Development*, 86(3), 844–863.
- Carranza, José Antonio, Carmen González-Salinas, and Ester Ato. 2013. A Longitudinal Study of Temperament Continuity through IBQ, TBAQ, and CBQ. *Infant Behavior and Development*, 36(4), 749–761. <<https://doi.org/10.1016/j.infbeh.2013.08.002>>
- Casalin, Sara, Patrick Luyten, Nicole Vliegen, and Patrick Meurs. 2012. The Structure and Stability of Temperament from Infancy to Toddlerhood: A One-Year Prospective Study. *Infant Behavior and Development*, 35(1), 94–108. <<https://doi.org/10.1016/j.infbeh.2011.08.004>>
- Cosentino-Rocha, Luciana, Vivian Caroline Klein, and Maria Beatriz Martins Linhares. 2014. Effects of Preterm Birth and Gender on Temperament and Behavior in Children. *Infant Behavior and Development*, 37(3), 446–456. <<https://doi.org/10.1016/j.infbeh.2014.04.003>>
- Dalecki, Michael and Fern K. Willits. 1991. Examining Change Using Regression Analysis: Three Approaches Compared. *Sociological Spectrum*, 11(2), 127–145.
- Desmarais, Eric, Mirjana Majdandžić, Maria A. Gartstein, David J. Bridgett, and Brian F. French. 2019. Cross-Cultural Differences in Temperament: Comparing Paternal Ratings of US and Dutch Infants. *European Journal of Developmental Psychology*, 16(2), 137–151.

- Denham, Susanne A., Kimberly A. Blair, Elizabeth DeMulder, Jennifer Levitas, Katherine Sawyer, Sharon Auerbach-Major, and Patrick Queenan. 2003. Preschool Emotional Competence: Pathway to Social Competence? *Child Development*, 74(1), 238–256. <<https://doi.org/10.1111/1467-8624.00533>>
- Eiden, Rina D., Ellen P. Edwards, and Kenneth E. Leonard. 2004. Predictors of Effortful Control among Children of Alcoholic and Nonalcoholic Fathers. *Journal of Studies on Alcohol*, 65(3), 309–319.
- Gartstein, Maria A., and Mary K. Rothbart. 2003. Studying Infant Temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and 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)>
- Gartstein, Maria A., Diane L. Putnick, Keumjoo Kwak, Chun-Shin Hahn, and Marc H. Bornstein. 2015. Stability of Temperament in South Korean Infants from 6 to 12 to 18 Months: Moderation by Age, Gender, and Birth Order. *Infant Behavior and Development*, 40, 103–107. <<https://doi.org/10.1016/j.infbeh.2015.05.002>>
- Goodman, Robert, Tamsin Ford, Helen Simmons, Rebecca Gatward, and Howard Meltzer. 2000. Using the Strengths and Difficulties Questionnaire (SDQ) to Screen for Child Psychiatric Disorders in a Community Sample. *The British Journal of Psychiatry*, 177(6), 534–539.
- Hitachi, Yurika, and Kaoru Nishimura. 2015. The Effects of Moving on Children: Implications for Social Support and Adaptation. *Waseda Journal of Clinical Psychology*, 15(1), 111–123.
- Joseph, Heather M., Nicole E. Lorenzo, Nadiyah Fisher, Danielle R. Novick, Cassandra Gibson, Scott D. Rothenberger, Jill E. Foust, and Andrea Chronis-Tuscano. 2023. Research Review: A Systematic Review and Meta-Analysis of Infant and Toddler Temperament as Predictors of Childhood Attention-Deficit/Hyperactivity Disorder. *Journal of Child Psychology and Psychiatry*, 64(5), 715–735. <<https://doi.org/10.1111/jcpp.13753>>
- Kanzaki, Mami, Shizuka Kawamoto, Asami Senoo, Yuki Nakata, Katsuki Higo, Sun Yi, Naoko Okamoto, Yuko Yasuda, Tatsuya Sato, Hanako Suzuki, and Yuko Yato. 2020. “Ibaraki Kohoto Chosa” ni okeru Rinriteki Hairyo (Practical and Ethical Considerations in the Ibaraki Cohort Study). *Japanese Journal of Human Sciences of Health-Social Services*, 26(2), 17–25. (in Japanese)
- Kanzaki, Mami, Asami Senoo, Yi Sun, Katsuki Higo, Teppei Tsuchimoto, Yuki Nakata, Hanako Suzuki, Yuko Yasuda, Tatsuya Sato, Naoko Okamoto, and Yuko Yato. 2022. Nani ga Chosa e no Sanka Keizoku o Sasaeru no ka: Ninpu o Taisho to Shita Judan Kenkyu “Ibaraki Kohoto Chosa” o Moto ni (What Supports Continued Participation in Surveys?: Based on the Ibaraki Cohort Study, a Longitudinal Study of Pregnant Women). *The Japanese Journal of Personality*, 30(3), 130–133. <<https://doi.org/10.2132/personality.30.3.2>> (in Japanese)
- Klein, Vivian Caroline, Luciana Cosentino Rocha, Francisco Eulógio Martinez, Samuel P. Putnam, and Maria Beatriz Martins Linhares. 2013. Temperament and Behavior Problems in Toddlers Born Preterm and Very Low Birth Weight. *The Spanish Journal of Psychology*, 16, E18. <<https://doi.org/10.1017/sjp.2013.30>>
- Komsi, Niina, Katri Räikkönen, Anu-Katriina Pesonen, Kati Heinonen, Pertti Keskivaara, Anna-Liisa Järvenpää, and Timo E. Strandberg. 2006. Continuity of Temperament from Infancy to Middle Childhood. *Infant Behavior and Development*, 29(4), 494–508.
- Kochanska, Grazyna, Kathleen T. Murray, and Elena T. Harlan. 2000. Effortful Control in Early Childhood: Continuity and Change, Antecedents, and Implications for Social Development. *Developmental Psychology*, 36(2), 220–232. <<https://doi.org/10.1037/0012-1649.36.2.220>>

- Lu, Xi, Zi Chen, Xiaoyi Cui, Masayo Uji, Wataru Miyazaki, Masako Oda, Toshiaki Nagata, Toshinori Kitamura, and Takahiko Katoh. 2012. Effects of Temperament and Character Profiles on State and Trait Depression and Anxiety: A Prospective Study of a Japanese Youth Population. *Depression Research and Treatment*, 2012, 604684. <<https://doi.org/10.1155/2012/604684>>
- Leaper, Campbell. 2002. Parenting Girls and Boys. In Marc H. Bornstein (ed.), *Handbook of Parenting: Children and Parenting* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 189–225.
- Matsuishi, Toyojiro, Miki Nagano, Yuko Araki, Yoshiyuki Tanaka, Mizue Iwasaki, Yushiro Yamashita, Shinichiro Nagamitsu, Chiho Iizuka, Takashi Ohya, Kunihiro Shibuya, Munetsugu Hara, Kentaro Matsuda, Akira Tsuda, and Tatsuyuki Kakuma. 2008. Scale Properties of the Japanese Version of the Strengths and Difficulties Questionnaire (SDQ): A Study of Infant and School Children in Community Samples. *Brain and Development*, 30(6), 410–415.
- Mattera, Jennifer A., Allegra X. Campagna, Sherryl H. Goodman, Maria A. Gartstein, Gregory R. Hancock, Zachary N. Stowe, D. Jeffrey Newport, and Bettina T. Knight. 2023. Associations between Mothers’ and Fathers’ Depression and Anxiety Prior to Birth and Infant Temperament Trajectories over the First Year of Life: Evidence from Diagnoses and Symptom Severity. *Journal of Affective Disorders*, 343, 31–41. <<https://doi.org/10.1016/j.jad.2023.09.023>>
- Miyoshi, Ryoei, Hisae Matsuo, Ryuichiro Takeda, Hiroyuki Komatsu, Hiroshi Abe, and Yasushi Ishida. 2016. Burnout in Japanese Residents and Its Associations with Temperament and Character. *Asian Journal of Psychiatry*, 24, 5–9.
- Murotsu, Fumiko, Yuko Takei, Masaharu Terasaki, and Masako Kadota. 2020. Sango 1-kagetsu to 3-kagetsu ji ni okeru Bonyu Ikuji Sutoresu no Eiyoho-betsu Hikaku (Comparison of Breastfeeding Stress at 1 and 3 months after Delivery by Nutrition Methods.). *Health Sciences and Human Formation*, 6(1), 33–43. <<https://doi.org/10.18883/johcu.0601.03>> (in Japanese)
- Nakagawa, Atsuko, and Masune Sukigara. 2005. Nyuji no Kodo no Kaishaku ni okeru Bunkasa wa IBQ-R Nihon-ban ni Dono Yoni Han’ei Sareru ka (How Are Cultural Differences in the Interpretation of Infant Behavior Reflected in the Japanese Revised Infant Behavior Questionnaire?). *Japanese Journal of Educational Psychology*, 53(4), 491–503. (in Japanese)
- Nakagawa, Atsuko, Yuka Kimura, Masune Sukigara, and Rie Mizuno. 2011. Nyuyoji no Kodo no Chekku Risuto (ECBQ) Tanshuku-ban (The Development of a Short Form of the Japanese Early Childhood Behavior Questionnaire). *Studies in Humanities and Cultures*, 16, 1-15. (in Japanese)
- Nakagawa, Atsuko, and Masune Sukigara. 2014. The Effects of Soothing Techniques and Rough-and-Tumble Play on the Early Development of Temperament: A Longitudinal Study of Infants. *Child Development Research*, 2014. <<https://doi.org/10.1155/2014/741373>>
- National Institute of Child Health and Human Development Early Child Care Research Network. 2003. Social Functioning in First Grade: Associations with Earlier Home and Child Care Predictors and with Current Classroom Experiences. *Child Development*, 74(6), 1639–1662. <<https://doi.org/10.1046/j.1467-8624.2003.00629.x>>
- Nigg, Joel T. 2006. Temperament and Developmental Psychopathology. *Journal of Child Psychology and Psychiatry*, 47(3-4), 395–422.
- Niu, Yubai, Lingyan Zhang, Zesheng Hao, and Yuzhu Ji. 2019. Relationship between Temperament, Parenting Style and Resilience of Children Aged 3–5 Years. *Journal of Zhejiang University (Medical Sciences)*, 48(1), 75–82. <<https://doi.org/10.3785/j.issn.1008-9292.2019.02.12>>
- Ogura, Tamiko. 2015. Nyuyojiki no Kishitsu no Ikkansei to Seisa (Consistency and Gender Differences of Temperament in Infancy and Early Childhood). *Journal of Contemporary Human Life Science*,

- Tezukayama University, 11, 65–74. (in Japanese)
- Ohtani, Noriko, Hatsumi Yamamoto, Tomomi Makino, Saeko Matsuoka, Akiko Uemura, Junko Tamaki, Noriko Yamakawa, Masaru Ido, and Masatoshi Kawai. 2006. Nihon no Kodomo no Sukoyakana Mirai no Tame ni: Dai 2 Ho: Sukusuku Kohoto ni okeru Mie Kenkyu Gurupu deno Tanki Yobiteki Kenkyu no Hikensha Kakuho to Sono Iji (For Strong and Healthy Future of Japanese Children: The Second Issue—Recruiting and Keeping Subjects in the Pilot Study of Sukusuku-cohort in Mie). *IRYO*, 60(9), 569–575. <<https://doi.org/10.11261/iryo1946.60.569>> (in Japanese)
- Putnam, Samuel P., Maria A. Gartstein, and Mary K. Rothbart. 2006. Measurement of Fine-Grained Aspects of Toddler Temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, 29(3), 386–401.
- Putnam, Samuel P., Mary K. Rothbart, and Maria A. Gartstein. 2008. Homotypic and Heterotypic Continuity of Fine-Grained Temperament during Infancy, Toddlerhood, and Early Childhood. *Infant and Child Development: An International Journal of Research and Practice*, 17(4), 387–405.
- Putnam, Samuel P., Julia F. Jacobs, Maria A. Gartstein, and Mary K. Rothbart. 2010. *Development and Assessment of Short and Very Short Forms of the Early Childhood Behavior Questionnaire*. Poster presented at the International Conference on Infant Studies, Baltimore, MD.
- Putnam, Samuel P., Amy L. Helbig, Maria A. Gartstein, Mary K. Rothbart, and Esther Leerkes. 2014. Development and Assessment of Short and Very Short Forms of the Infant Behavior Questionnaire Revised. *Journal of Personality Assessment*, 96(4), 445–458.
- Rothbart, Mary K., and John E. Bates. 2007. Temperament. In William Damon, Richard M. Lerner, and Nancy Eisenberg (eds.), *Handbook of Child Psychology*, vol. 3, 6th ed. New York: Wiley, 99–166. <<https://doi.org/10.1002/9780470147658.chpsy0303>>
- Rothbart, Mary K., Brad E. Sheese, M. Rosario Rueda, and Michael I. Posner. 2011. Developing Mechanisms of Self-Regulation in Early Life. *Emotion Review*, 3(2), 207–213.
- Rothbaum, Fred, John R. Weisz, Martha Pott, Kazuo Miyake, and Gail Morelli. 2000. Attachment and Culture: Security in the United States and Japan. *American Psychologist*, 55(10), 1093–1104. <<https://doi.org/10.1037/0003-066X.55.10.1093>>
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.
- Ruddy, Margaret G., and Marc H. Bornstein. 1982. Cognitive Correlates of Infant Attention and Maternal Stimulation over the First Year of Life. *Child Development*, 53(1), 183–188.
- Sieber, Franziska, and Norbert Zmyj. 2022. Stability and Structure of Infant and Toddler Temperament in Two Longitudinal Studies in Germany. *Infant Behavior and Development*, 67, 101714. <<https://doi.org/10.1016/j.infbeh.2022.101714>>
- Slagt, Meike, Judith Semon Dubas, Maja Deković, and Marcel A. G. van Aken. 2016. Differences in Sensitivity to Parenting Depending on Child Temperament: A Meta-Analysis. *Psychological Bulletin*, 142(10), 1068–1110. <<https://doi.org/10.1037/bul0000061>>
- Sroufe, L. Alan. 1977. Wariness of Strangers and the Study of Infant Development. *Child Development*, 48(3), 731–746.
- Stern, Daniel N. 2018. *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. London: Routledge.
- Thomas, Alexander, and Stella Chess. 1977. *Temperament and Development*. New York: Brunner/Mazel.

- Trommsdorff, Gisela, and Wolfgang Friedlmeier. 2010. Preschool Girls' Distress and Mothers' Sensitivity in Japan and Germany. *European Journal of Developmental Psychology*, 7(3), 350–370.
- Vroman, Lisa N., Sharon L. Lo, and C. Emily Durbin. 2014. Structure and Convergent Validity of Children's Temperament Traits as Assessed by Experimenter Ratings of Child Behavior. *Journal of Research in Personality*, 52, 6–12.