

Article

The Boston University Twin Project (BUTP) II

Kimberly J. Saudino¹ and Jody M. Ganiban²

¹Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA and ²Department of Psychology, The George Washington University, Washington, DC, USA

Abstract

The Boston University Twin Project (BUTP) uses a multimethod, longitudinal approach to study the role of genetic and environmental factors on the development of child temperament and related behaviors in early childhood. There are two phases in this project. The first, described in the previous *Twin Research and Human Genetics* special issue on twin registries, focused on activity level and comprised over 300 twin pairs assessed in the home and laboratory at ages 2 and 3. In this article, we describe subject recruitment, sample characteristics, and study procedures and measures of the second phase of the BUTP. This recent study focuses more broadly on the development of multiple temperament dimensions and explores associations between temperament trajectories, parenting and child adjustment in a new cohort of approximately 300 twin pairs assessed at 3, 4 and 5 years of age.

Keywords: Temperament; parenting; longitudinal; multimethod; growth

(Received 30 May 2019; accepted 13 June 2019; First Published online 13 September 2019)

Background

The initial phase of the Boston University Twin Project (BUTP) was a multisituation, multimethod, longitudinal investigation of genetic influences on the temperament dimension of activity level and related behaviors in early childhood. The sample and methods were described in Saudino and Asherson (2013). Our use of multiple measures of activity level (actigraphs, parent ratings and observer ratings) in this early BUTP study has provided unique evidence of situation-specific (Saudino & Zapfe, 2008), measure-specific (Saudino, 2009) and age-specific (Saudino, 2012) genetic influences on activity level; and genetic links between activity level and hyperactivity (Ilott et al., 2010; Ilott, Saudino, & Asherson, 2010), attention problems (Saudino et al., 2018) and shyness (Frazier-Wood & Saudino, 2017). We have also explored genetic and environmental contributions to a number of novel phenotypes in early childhood including elicited imitation (Fenstermacher & Saudino, 2007), inhibitory control (Gagne & Saudino, 2016), callous/unemotional behaviors (Flom & Saudino, 2017), autistic-like traits (Edelson & Saudino, 2009) and positive affect (Flom et al., 2018). Although data collection for this initial sample was completed in 2007, the dataset is still yielding novel contributions to the literature that relate to child activity and behavioral outcomes (e.g., Flom et al., 2019).

Here, we introduce phase II of the BUTP, which involves a new preschool sample longitudinally assessed at ages 3, 4 and 5 years. The second phase of the BUTP builds on our prior work but has a broader focus and answers very different developmental questions. While we still include measures of activity level in this study, our emphasis is on the development of temperament more broadly and

genetic and environmental contributions to *growth* in temperament. We are interested in understanding the factors that underlie variation in developmental trajectories of both, temperament and parenting, and their links with developmental outcomes. Data collection for this study began in 2012 and ended in 2018. This study has yielded a rich dataset that includes multiple dimensions of child temperament and positive and negative emotional/behavioral outcomes, along with parenting assessed at three time points, using a multimethod approach. Data analyses exploring genetic and environmental contributions to growth in these domains are underway. In this article, we provide a detailed description of the sample, study procedures and measures in the hopes of sparking future collaborations.

Recruitment and Sample Characteristics

As with the first phase of the BUTP, twins were recruited from birth records supplied by the Massachusetts Registry of Vital Records. Twins with birth weights less than 1750 g, gestational age less than 34 weeks or with known developmental or health problems (e.g., autism, Down syndrome) were excluded. The average age of mother at time of birth was 34.4 years (range 20.5–48.4). Forty-seven percent of the mothers had some form of fertility treatment, which likely reflects the fact that Massachusetts is the state with the highest rate of assisted reproductive technology births, and is consistent with the demographics of our sample (Pew Research, 2018).

Sample. Table 1 summarizes the twin sample at each age. All twins within a pair were the same sex; this ensured that sex differences between dizygotic twin siblings did not contribute to behavioral differences between siblings. Three hundred and ten pairs of twins participated in the age 3 laboratory assessments; of these, 286 pairs (92.3%) were assessed again at age 4 and

Author for correspondence: Kimberly J. Saudino, Email: ksaudino@bu.edu

Cite this article: Saudino KJ and Ganiban JM. (2019) The Boston University Twin Project (BUTP) II. *Twin Research and Human Genetics* 22: 691–694, <https://doi.org/10.1017/thg.2019.58>

© The Author(s) 2019.

Table 1. Number of twin pairs seen in the laboratory, by zygosity and sex, at each age

	Age 3	Age 4	Age 5
<i>N</i> twin pairs	310	286 ^a	274 ^b
MZ males	63	61	58
MZ females	60	57	58
DZ males	90	83	75
DZ females	97	85	83
Dates of births	09/2009–05/2013	09/2009–05/2013	09/2009–05/2013
Dates of visits	10/2012–05/2017	10/2013–05/2017	10/2014–05/2018

MZ=Monozygotic; DZ=Dizygotic.

^a*N* = 291 pairs including five families with questionnaire data, but no age 4 laboratory visit.

^b*N* = 281 pairs including seven families with questionnaire data, but no age 5 laboratory visit.

274 pairs (88.4%) at age 5. Five of the families who did not return at age 4 and 7 of those who did not return at age 5 completed questionnaire data, thus we have some longitudinal data on 291 pairs at age 4 and 281 pairs at age 5. Although the sample was predominately Caucasian (89.6%), ethnicity was generally representative of the Massachusetts population (1.6% Black, 1.9% Asian, 6.2% Mixed and 6.2% Hispanic or Latino). The parents were highly educated, with over 50% of the primary caregivers having a bachelor's degree or higher. Socioeconomic status (SES) was primarily middle to upper-middle class, but ranged from low to high SES.

Study Procedure

Overview. Twins and one primary caretaker (95% mothers) visited the BUTP laboratory within approximately 1 month of the twins' 3rd, 4th and 5th birthdays. Each assessment lasted 2½–3 h, during which the twins participated in a number of structured situations designed to assess multiple facets of child temperament and parent–child interactions. Observational measures and standardized tests assessing cognitive abilities, preschool readiness and prosocial behavior were also obtained. Tasks were arranged into four blocks and organized to minimize cognitive fatigue and/or negative affective carryover from one block to the next. Blocks were counterbalanced across first- and second-born twins. All assessments were video recorded for later behavioral coding. Within a twin pair, twins were individually assessed by different testers and behavioral ratings from video recordings were made by different coders. In addition to our behavioral assessments, at all ages, parents completed a battery of questionnaires designed to inform about child temperament, behavior problems, family characteristics and demographics. All procedures were approved by the Boston University Institutional Review Board, and primary caregivers provided informed consent.

Zygosity. At age 3, cheek scrapings were used to obtain DNA samples from twins. DNA extraction was performed at the Institute of Psychiatry (London, UK). Zygosity was determined via DNA analyses by genotyping 10 highly polymorphic simple sequence repeat markers in each member of a twin pair. For 10 families who declined to provide DNA samples, zygosity was determined using parents' responses to physical similarity questionnaires.

Feedback to participants. With the exception of information about twins' zygosity based on DNA analyses, families do not

receive specific information about their children. However, the BUTP publishes annual newsletters, which are sent to all families who have participated in any of our research projects. The newsletter's intent is to disseminate to a general (i.e., nonscientific) audience new information regarding our research findings.

Measures

Unless otherwise noted, the same measures were administered to twins at each age. This ensured that any observed changes in behavior were not due to methodological differences across age. Measures marked with an asterisk were also included in our Phase I sample and allow the possible combination of data across samples.

Laboratory-assessed temperament

The Laboratory Temperament Assessment Battery — Preschool Version (Lab-TAB; Goldsmith et al., 1995) was used to assess temperament within standardized and structured situations. The Lab-TAB Fear/Anger episodes (Stranger Approach and Imperfect Circles), Exuberance episodes (Popping Bubbles and Surprise), Activity episodes* (Corral of Balls, Arc of Toys and Fidgeting Video) and Interest/Persistence episodes (Bead Sorting and Coffee Pot) were used to elicit specific temperament behaviors. For each of these episodes, trained observers coded the video-recorded data for the dimensions of *negative affect*, *positive affect*, *attention*, *persistence* and *social engagement* using a global five-point behavioral rating of each based on the Bayley Behavior Rating Scale (Bayley, 2006). Summary scores for each dimension were obtained by averaging the behavioral ratings across the nine Lab-TAB episodes. *Activity level* was assessed with Minimitter actigraphs*, attached one per limb by means of Tyvek wristbands. A composite activity score was formed based on the mean of the four limb scores. A laboratory-based measure of *inhibitory control* was also obtained using the Flanker Test from the NIH Toolbox: Early Childhood Cognitive Battery (Zelazo et al., 2013; see 'Cognitive abilities' below).

Tester-rated temperament

Following the laboratory visits, testers of each twin completed the Infant Behavior Record* (IBR; Bayley, 1969) to obtain behavioral ratings of temperament based on behaviors observed across the entire laboratory visit (i.e., including cognitive testing and other non-Lab-TAB activities). Factor analysis of the IBR has yielded three temperament dimensions: Activity, Affect/Extraversion and Task Orientation (Matheny, 1983). Because the IBR is a frequently used observer-rated measure of temperament in behavioral genetic research, including the earlier BUTP sample (e.g., Frazier-Wood & Saudino, 2017), the inclusion of this measure in the present study allows us to compare our data with previous findings and will help to address issues of replication.

Parent reports of temperament

Parents rated the temperament characteristics of each twin with the Child Behavior Questionnaire — Short Form (CBQ-SF; Putnam & Rothbart, 2006). The CBQ-SF assesses 15 dimensions of temperament, including Positive Anticipation, High-Intensity Pleasure, Smiling/Laughter, Activity Level, Impulsivity, Shyness, Discomfort, Fear, Anger/Frustration, Sadness, Soothability, Inhibitory Control, Approach, Attentional Focusing, Low-Intensity Pleasure and Perceptual Sensitivity, as well as three superfactors, Surgency, Negative Affectivity and Effortful Control.

Observed parent-child interaction

The primary caretaker was video recorded while separately interacting with each twin during an Etch-A-Sketch drawing task, a free play session, and during clean-up (approximately 10 min total). These parent-child interaction tasks have been widely used in studies in early and middle childhood (e.g., NICHD Study of Early Child Care and Youth Development). Observations were coded for parent behaviors (positive control, negative control, positive affect, negative affect and responsiveness), child behaviors (positive affect, negative affect, responsiveness, compliance, autonomy, on-task behavior and activity level) and dyadic interaction (reciprocity, conflict and cooperation) using the Parent-Child Interaction System (Deater-Deckard et al., 1997).

Parent reports of parenting behaviors

Parent reports of parenting were based on measures used in the Twins Early Development Study at similar ages (see Knafo & Plomin, 2006). Parent positive and negative affects toward each twin were assessed using the Parent Feelings Questionnaire* (Deater-Deckard, 1996). A measure of harsh discipline* was obtained via a widely used semistructured interview modified to a parent-report format (Deater-Deckard, 2000; Deater Deckard et al., 1996). For each twin, parents rated the frequency of use for a variety of discipline strategies, yielding a child-specific global rating of harshness of discipline.

Behavior problems

Parents reported on their twins' behavior problems using the Child Behavior Checklist for Ages 1½–5* (CBCL; Achenbach & Rescorla, 2000). In addition to the traditional CBCL scoring yielding three higher order scales (Internalizing, Externalizing and Total Behavior problems), seven syndrome scales (Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Attention Problems, Aggressive Behavior and Sleep Problems) and seven DSM-oriented scales (Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention-Deficit/Hyperactivity Problems, Stress Problems, Autism Spectrum Problems and Oppositional Defiant Problems), we included scales of Callous/unemotional behaviors (Willoughby et al., 2011) and Irritability (Wiggins et al., 2014). Parents also reported twins' behavior problems on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), which yields information on emotional symptoms, conduct problems, hyperactivity/inattention and peer relationship problems.

Prosocial behaviors

Parent ratings of twins' prosocial behaviors (e.g., shares, considerate, kind, caring) were obtained via the Prosocial subscale of the SDQ. A *sharing task* provided an observational measure of prosocial behavior. This task, widely used in studies of child prosocial behavior (e.g., Blake & Rand, 2010), is a child version of the dictator game where children are presented with 10 stickers and told that they can do whatever they want with them, keep them all or give some or all of the stickers to their twin. The number of stickers given to their co-twin indexed sharing.

Academic readiness

The Bracken School Readiness Assessment — Third Edition (Bracken, 2007) provided a standardized measure of academic

readiness. This test assesses knowledge of color, letters, numbers/counting, sizes, comparisons and shapes in children from 3 to 7 years and is a good predictor of student outcomes (Panter & Bracken, 2009).

Cognitive abilities

The NIH Toolbox Early Childhood Cognitive Battery (Zelazo et al., 2013) was used to assess executive functioning (inhibitory control and set shifting), receptive vocabulary and episodic memory. This battery, recommended for ages 3–6, is a series of computerized game-like tasks that include the Flanker, Dimensional Change Card Sort, Picture Vocabulary and Picture Sequence Memory subtests. The Flanker task assesses inhibitory control and attention by asking the child to focus on a target stimulus while inhibiting attention to stimuli flanking it. The Dimensional Change Card Sort is a set-shifting task that requires children to match a series of bivalent test pictures (e.g., yellow balls and blue trucks) to target pictures, first according to one dimension (e.g., color) and then, after a number of trials, according to the other dimension (e.g., shape). In the Picture Vocabulary task (receptive vocabulary), children were presented with a recording of a word and four photographic images on the computer screen and asked to select the picture that most closely matched the meaning of the word (Gershon et al., 2013). In the Picture Sequencing task, children were shown an arbitrary ordering of pictures and asked to reproduce the sequence (Bauer et al., 2013).

Household chaos

Parent perceptions of environmental confusion in the home were obtained using the Confusion, Hubbub, and Order Scale (Matheny et al., 1995). This brief measure assesses the degree of organization and calmness in the twins' home.

Height* and weight*

Children were weighed on a digital scale and measured with a stadiometer wearing light indoor clothing and no shoes.

Hair cortisol (age 5 only)

Testers, using sterilized hair scissors, cut 30 mg of hair from each child's posterior vertex as close as possible to the scalp. The 3 cm closest to the scalp was assayed for cortisol levels at the University of Massachusetts, Amherst. Human scalp hair grows at a rate of approximately 1 cm/month, so the 3-cm sample serves as an index of chronic cortisol output over the past 3 months.

Demographics*

Parents completed a demographic questionnaire regarding race/ethnicity, family composition, parent education and occupation, pregnancy and birth, and twins' physical similarities, health and daycare.

Conclusions

The second phase of the BUTP comprises 1740 individual comprehensive laboratory-based assessments and approximately 5000 h of behavioral observations and has yielded a vast amount of data. This unique dataset allows us to address important questions regarding the implications and etiology of developmental change in temperament in preschoolers. We are currently planning to follow up this second cohort of twins pending funding.

Financial support. The Boston University Twin Project (BUTP) is supported by grants MH062375 and HD068435 to Dr. Saudino.

References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms and Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Bauer, P. J., Dikmen, S. S., Heaton, R. K., Mungas, D., Slotkin, J., & Beaumont, J. L. (2013). III. NIH toolbox cognition battery (CB): Measuring episodic memory. *Monographs of the Society for Research in Child Development*, 78, 34–48.
- Bayley, N. (1969). *Manual for the Bayley scales of infant development*. New York: The Psychological Corporation.
- Bayley, N. (2006). *Bayley scales of infant and toddler development technical manual* (3rd ed.). San Antonio, TX: PsychCorp.
- Blake, P. R., & Rand, D. G. (2010). Currency value moderates equity preference among young children. *Evolution and Human Behavior*, 31, 210–221.
- Bracken, B. A. (2007). *Bracken school readiness assessment* (3rd ed.). Examiner's manual. San Antonio, TX: Pearson.
- Deater-Deckard, K. (1996). *The parent feelings questionnaire*. London, UK: Institute of Psychiatry.
- Deater-Deckard, K. (2000). Parenting and child behavioral adjustment in early childhood: A quantitative genetic approach to studying family processes. *Child Development*, 71, 468–484.
- Deater-Deckard, K., Dodge, K., Bates, J. E., & Pettit, G. (1996). Physical discipline among African-American and European-American mothers: Links to children's externalizing behaviors. *Developmental Psychology*, 32, 1065–1072.
- Deater-Deckard, K., Pylas, M., & Petrill, S. A. (1997). *Parent-Child Interaction Coding System*. London, UK: Institute of Psychiatry.
- Edelson, L. R., & Saudino, K. J. (2009). Genetic and environmental influences on autistic-like behaviors in 2-year-old twins. *Behavior Genetics*, 39, 255–264.
- Fenstermacher, S. K., & Saudino, K. J. (2007). Toddler see, toddler do? Genetic and environmental influences on laboratory-assessed elicited imitation. *Behavior Genetics*, 37, 639–647.
- Flom, M., & Saudino, K. J. (2017). Callous-unemotional behaviors in early childhood: Genetic and environmental contributions to stability and change. *Development and Psychopathology*, 29, 1227–1234.
- Flom, M., Wang, M., Uccello, K. J., & Saudino, K. J. (2018). Parent- and observer-rated positive affect in early childhood: Genetic overlap and environmental specificity. *Behavior Genetics*, 48, 432–439.
- Flom, M., White, D., Ganiban, J., & Saudino, K. J. (2019). Longitudinal links between callous-unemotional behaviors and parenting in early childhood: A genetically informed design. *Journal of the American Academy of Child and Adolescent Psychiatry*. Advance online publication. doi:10.1016/j.jaac.2019.03.013
- Frazier-Wood, A., & Saudino, K. J. (2017). Activity level in the lab: Overlap with shyness indicates it is more than pure motoric activity. *Developmental Psychology*, 53, 1611–1619.
- Gagne, J. R., & Saudino, K. J. (2016). The development of inhibitory control in early childhood: A twin study from 2–3 years. *Developmental Psychology*, 52, 391–399.
- Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., . . . Weintraub, S. (2013). IV. NIH toolbox cognition battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*, 78, 49–69.
- Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1995). *The laboratory temperament assessment battery – preschool version: Description of procedures*. Unpublished manuscript, University of Wisconsin-Madison.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology & Psychiatry*, 38, 581–586.
- Ilott, N., Saudino, K. J., & Asherson, P. (2010). Genetic influences on attention deficit hyperactivity disorder symptoms from age 2 to 3: A quantitative and molecular genetic investigation. *BMC Psychiatry*, 10, 102.
- Ilott, N., Saudino, K. J., Wood, A., & Asherson, P. (2010). A genetic study of ADHD and activity level in infancy. *Genes, Brain, and Behavior*, 9, 296–304.
- Knafo, A., & Plomin, R. (2006). Parental discipline and affection, and children's prosocial behavior: Genetic and environmental links. *Journal of Personality and Social Psychology*, 90, 147–164.
- Matheny, A. P. (1983). A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. *Child Development*, 54, 356–360.
- Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of Applied Developmental Psychology*, 16, 429–444.
- Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken school readiness assessment for predicting first grade readiness. *Psychology in the Schools*, 46, 397–409.
- Pew Research. (2018). *U.S. births via assisted reproductive technology up more than threefold since 1996*. Washington, DC: Author.
- Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment*, 87, 102–112.
- Saudino, K. J. (2009). Do different measures tap the same genetic influences? A multi-method study of activity level in young twins. *Developmental Science*, 12, 626–633.
- Saudino, K. J. (2012). Sources of continuity and change in activity level in early childhood. *Child Development*, 83, 266–281.
- Saudino, K. J., & Asherson, P. (2013). The Boston University Twin Project (BUTP). *Twin Research and Human Genetics*, 16, 449–450.
- Saudino, K. J., Wang, M., Flom, M., & Asherson, P. (2018). Genetic and environmental links between motor activity level and attention problems in early childhood. *Developmental Science*, 21, e12630.
- Saudino, K. J., & Zapfe, J. A. (2008). Genetic influences on activity level in early childhood: Do situations matter? *Child Development*, 79, 930–943.
- Wiggins, J. L., Mitchell, C., Stringaris, A., & Leibenluft, E. (2014). Developmental trajectories of irritability and bidirectional associations with maternal depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53, 1191–1205.
- Willoughby, M. T., Waschbusch, D. A., Moore, G. A., & Propper, C. B. (2011). Using the ASEBA to screen for callous unemotional traits in early childhood: Factor structure, temporal stability, and utility. *Journal of Psychopathology and Behavioral Assessment*, 33, 19–30.
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. HIATRNH toolbox cognition battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78, 16–33.