Published in final edited form as:

J Child Psychol Psychiatry. 2017 May ; 58(5): 564–572. doi:10.1111/jcpp.12670.

Tipping Points? Curvilinear Associations Between Activity Level and Mental Development in Toddlers.

Megan Flom, Madeleine Cohen, Kimberly J. Saudino

Boston University, 64 Cummington Mall, Boston, MA 02215

Abstract

Background—The Theory of Optimal Stimulation (Zentall & Zentall, 1983) posits that the relation between activity level (AL) and cognitive performance follows an inverted U shape where mid-range AL predicts better cognitive performance than AL at the extremes.

Methods—We explored this by fitting linear and quadratic models predicting mental development from AL assessed via multiple methods (parent ratings, observations, and actigraphs) and across multiple situations (lab play, lab test, home) in over 600 2- and 3-year-old twins.

Results—Only observed AL in the lab was curvilinearly related to mental development scores. Results replicated across situations, age, and twin samples, providing strong support for the optimal stimulation model for this measure of AL in early childhood.

Conclusions—Different measures of AL provide different information. Observations of AL which include both qualitative and quantitative aspects of AL within structured situations are able to capture beneficial aspects of normative AL as well as detriments of both low and high AL.

Keywords

Activity level; early childhood; curvilinear; mental development; optimal stimulation model

Introduction

There is an abundance of evidence showing that high activity level (AL) in childhood is associated with a variety of non-optimal outcomes. This is true whether viewed as a temperament trait or as a clinically-relevant behavior (e.g., hyperactivity). High AL in early childhood is related to increased behavior problems (e.g., Fagot & O'Brien, 1994; Rothbart, Derryberry, & Hershey, 2000) more conflict and negativity in social interactions (Rothbart, Derryberry, & Posner, 1994), and lower academic achievement and IQ (Halverson & Waldrop, 1976; Martin, Olejnik, & Gaddis, 1994). It is also a key component of attention deficit/hyperactivity disorder (ADHD) and high levels of motor activity are included in 4 of the 9 symptoms in the hyperactive/impulsivity category for ADHD in the DSM-V. This research, however, has applied linear models to look at AL and its outcomes. While these studies clearly show that high AL is problematic, it is possible that low AL may also be

problematic. This will go undetected in analyses that look for simple linear associations and do not assess the entire spectrum of activity.

Although not well studied, there are hints that low AL may be problematic. Children described as having a “sluggish cognitive tempo” (SCT), a cognitive profile wherein the individual has a tendency to be “in a fog”, drowsy, daydream, and have slowed behavior or thinking (Becker, 2013), also have low energy and AL. A clinical case study of SCT finds negative cognitive, emotional and social developmental outcomes (Becker et al., 2016), suggesting that extremely low levels of AL and arousal are troublesome to child development. The potential curvilinear function between AL and developmental outcomes, however, has not been explored in a normative sample.

Though there is more focus on the negative implications of high AL, several important models suggest that AL has a beneficial function, which by implication, would mean that low AL is detrimental. For example, the *Optimal Stimulation Model* posits a curvilinear relation between AL and task performance such that a moderate amount of AL is associated with the best task performance (Zentall & Zentall, 1983). According to this model, it may be that the child’s AL is a behavioral representation of their arousal or level of stimulation to the environment, where children with high AL are under-stimulated and those with low AL are over-stimulated. Another possibility is that children with moderate AL are those who are better able to adjust their AL and arousal to the task at hand. Thus, AL may not be a behavioral representation of arousal level, but an indication of the child’s ability to moderate stimulation. Either way, this model proposes a purposeful conceptualization of childhood AL. The *Functional Immaturity Model* (Bjorklund, 1997), while not specific to AL, also provides support for the adaptive potential of childhood AL. This model points out that by maintaining a progressive view of childhood development whereby immature behaviors lead to “better” mature behaviors, we ignore their possible utility in childhood. For example, with respect to AL, it has been suggested that motor activity in infants and children allows for increased exploration and more concrete and varied interactions with the environment that facilitate learning (Eaton, McKeen, & Campbell, 2001; Piaget & Cook, 1952). Although not explicitly suggesting an “optimal” level, these ideas indicate the potential utility of AL and provide an important theoretical foundation for exploring AL in typical and atypical children. The *Evolutionary Trade-off Hypothesis* (Nettle, 2006) predicts better outcomes for individuals in the midrange on personality traits because it permits a more favorable trade-off between advantageous and disadvantageous consequences. As such, either extreme on a personality continuum would be more disadvantageous. Lastly, the *Functional Working Memory Model* (Rapport et al., 2009) proposes that AL in ADHD serves the purpose of augmenting prefrontal cortical hypoactivation during academic tasks demanding working memory. In other words, increased AL is a tool used by children with ADHD to improve their working memory performance.

All four models highlight the potential utility of AL but have been applied almost exclusively to clinical samples. Typical populations, when used, often serve as a comparison to atypical development (i.e., ADHD), with a focus on mean level differences rather than individual differences or patterns that are consistent across both groups. Furthermore, treating non-hyperactive children as a single group ignores meaningful variation in the

moderate to low range of AL (Fair, Bathula, Nikolas, & Nigg, 2012). It has been argued that the recent emphasis on the heterogeneity of ADHD should be applied to the entire range of childhood AL (Fair et al., 2012). Consistent with the emerging view of psychopathology from a more dimensional perspective, it is important to understand childhood AL across the entire distribution.

The limited research exploring the possible utility of AL suggests that moderate levels may be adaptive for cognitive development. In early infancy, AL is positively associated with cognitive development (Fagen, Singer, Ohr, & Fleckenstein, 1987; Matheny, 1989), but results are more mixed in older children (Maccoby, Dowley, Hagen, & Degerman, 1965). The inconsistent findings in older children may be the result of a curvilinear effect that has been masked by statistical approaches limited to exploring linear associations. A more consistent positive association between motor inhibition, the ability to regulate AL, and intellectual ability has emerged in community samples (Loo & Wenar, 1971; Maccoby et al., 1965). Furthermore, even though typically-developing children have lower mean levels of AL than children with ADHD, both typically-developing children and children with ADHD increase AL on cognitive tasks requiring working memory (Rapport et al., 2009) thus demonstrating the potential functional utility of AL in typical and atypical populations. Similarly, physical activity during math practice has been found to increase performance on an executive function task (Vazou & Smiley-Oyen, 2014). These findings of a beneficial impact of AL on cognitive performance would fit with the notion of an optimal level of AL.

Given the dearth of research on the potential detrimental developmental significance of both high *and* low AL, the present study is the first to explore both linear and curvilinear relations between AL and mental development across the transition from infancy to early childhood. Observer, parent, and mechanical measures of AL were used at ages 2 and 3, allowing us to examine the role of AL using different methodological approaches. Based on previous research we expected a negative linear association between AL and mental development, but we also predicted that there would be a curvilinear association when tested with the appropriate model. Because prior research has not explored this question, we had no apriori hypothesis regarding the relative amounts of variance in mental development explained by linear and quadratic components.

Methods

Sample

The Boston University Twin Project sample was recruited from birth records supplied by the Massachusetts Registry of Vital Records. Twins were selected preferentially for higher birth weight and gestational age. No twins with birth weights below 1750 grams or with gestational ages less than 34 weeks were included in the study. Twins were also excluded if they had a health problem that might affect motor activity (e.g., club foot) or had chromosomal abnormalities. Six hundred and twenty-six twins (313 pairs) participated in the age 2 assessments, and 608 twins returned for the age 3 assessments (96.8% retention rate). Race was generally representative of the Massachusetts population (85.4% White, 3.2% Black, 2% Asian, 7.3% Mixed, 2.2% Other). Socioeconomic status according to the

Hollingshead Four Factor Index ranged from low to upper middle class (range=20.5–66; $M=50.9$, $SD=14.1$).

Procedure Overview

Participants were assessed within approximately 2 weeks of their second and third birthdays. At each age, the procedure consisted of two 1-hour visits, 48-hours apart, to the laboratory. At the initial visit, informed consent from parents was obtained. One twin was assessed within a standardized test situation, while the other twin was assessed within a laboratory play situation. At the second visit, situations were reversed for each twin. The order of situations was counterbalanced across first- and second-born twins. The test situation involved administration of the Mental Scale of *Bayley Scales of Infant Development—Second Edition* (Bayley, 1993). The play situation comprised activity and inhibitory control episodes (arc of toys, corral of balls, workbench, fidget video, dinky toys, snack delay, gift) from the *Laboratory Temperament Assessment Battery—Preschool Version* (Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995). Twins were assessed by different testers; however, within age for each twin the tester was the same across the two laboratory situations.

Measures

Observer-rated Activity Level.—The *Infant Behavior Record* (IBR, Bayley, 1969) provided an observational measure of AL in the laboratory. Of the 30 items on the IBR, 25 evaluate broad dimensions of infant behavior, including interpersonal, affective, motivational, and sensory behavioral domains and are rated on a 5- or 9-point scale. Factor analyses of these 25 items (Matheny, 1980) has yielded an activity factor as well as two other temperament factors (task orientation and affect-extraversion). The activity factor, which was used in this study, includes the child's general level of body motion and degree of energy exhibited during the test situation. Although initially designed for children up to 30 months of age, the IBR has been successfully used to provide an observational measure of temperament in children 3 to 8 years of age (Deater-Deckard, Petrill, & Thompson, 2007; Saudino & Cherny, 2001). Following each visit the tester rated the child's activity level. The standardized unweighted items were aggregated to form composite observed AL scores for both the test and play situations. Internal consistencies, as indexed by Cronbach's alpha, were high (age 2 =.85; age 3=.84). Interrater agreement was adequate (age 2 =.60; age 3=.64).

Mechanical assessment of AL.—Twins wore Minimitter Acticals (actigraphs) for 48-hours in the home and the two lab situations. Four randomly-selected actigraphs were assigned to each twin, one for each limb and were attached to wrists and ankles by means of tyvek® adhesive or plastic wristbands (see Saudino & Zapfe, 2008 for a description of actical methodology). To adjust for variations in the total time that each instrument was worn in each situation, the number of activity units was converted to a rate per minute real time. Arm and leg activity counts were highly correlated (Age 2 $r=.69$, $p<.001$; Age 3 $r=.73$, $p<.001$), therefore, composite actigraph scores reflecting overall motor activity were calculated by averaging the four limb actigraph scores. Home AL reflects all activity outside the lab during the 2-day data collection period.

Parent-rated Activity Level.—The *Toddler Behavior Assessment Questionnaire* (TBAQ, Goldsmith, 1996) provided a parent-rating measure of AL as a temperament trait. The activity subscale of the TBAQ consists of 10 questions regarding the child's behavior in specific situations (e.g., "When playing inside the house, how often did your child climb over furniture?"). Parents (94% mothers) were asked to indicate on a 7-point scale how frequently the child demonstrated the behavior during the previous month (1=Never to 7=Always). The TBAQ is a widely-used assessment of different domains of temperament in early childhood, and has been proven to be a reliable and valid measure of toddler temperament (Goldsmith, 1996). Internal consistency for this scale was .78 and .79 at ages 2 and 3, respectively.

Mental Development.—The *Bayley Scales of Infant Development-Second Edition* (BSID-II; Bayley, 1993) was administered at both ages to obtain a standardized assessment of mental development. This measure is a well-established reliable and valid measure of cognitive ability that involves a series of interactive tasks designed to assess children's cognitive development in a variety of domains including vocabulary, spatial development, and memory. The BSID-II was standardized on a sample of 1700 children ranging in age from 1 to 42 months (Bayley, 1993). The mental scale provides a standardized Mental Development Index (MDI) score scaled to a mean of 100 ($SD=15$) reflecting general mental ability.

Statistical Analyses

All analyses were conducted using STATA 10 (StataCorp, 2007). Nested linear and quadratic models were fit to the data for each measure of AL at each age. Robust regression models using the CLUSTER option accounted for non-independence in the data arising from twins nested in families. R -squared change (R^2) was examined to determine if adding a quadratic component to the data significantly increased the amount of variance explained in mental development. In addition to examining whether the effects replicated across situation and age, one twin from each pair was randomly assigned to one of two groups (Random 1 or Random 2) to provide a within-family replication (these results are presented in the appendix). Because there was very little missing data within and across age complete cases analysis (i.e., listwise deletion) was used to handle the missing data (Graham, 2012).

Results

Descriptive and Correlational Statistics

Table 1 presents means for all variables. At both ages, the mean for mental development is within the normal range; however, as is typical with young twins (Plomin, DeFries, Knopik, & Neiderheiser, 2013), the mean at age 2 is at the lower end of the normal range but by age 3 is close to the standardization average. Although there are no standardized values for our AL measures, prior research suggests that twins do not differ from singletons in motor development (Brouwer, van Beijsterveldt, Bartels, Hudziak, & Boomsma, 2006).

Table 2 presents the intercorrelations between variables at each age. Within the lab, there was substantial agreement between observer-rated and actigraph AL; however, parent-rated

and actigraph AL in the home were only modestly correlated. AL in the test and play situations were highly correlated whether assessed by observers or actigraphs. Parent ratings of AL were weakly correlated with all measures of AL in the lab. At age 2 there were modest negative associations between the MDI and all measures of AL, but at age 3 the correlations between the MDI and actigraph AL in the play and home situations were nonsignificant. A similar pattern emerged when the sample was split by twins (Tables A.4, A.5).

Model-Fitting

Linear and quadratic models were first run separately for males and females; however, both sexes showed a similar pattern of results. Thus, because reporting results separately by sex would double the number of analyses without providing additional information, models were fit to the entire sample using AL scores residualized for sex thereby removing any variance due to sex. Table 3 presents the regression coefficients for the full quadratic model and summarizes the linear R^2 and R^2 after including the curvilinear component (standardized coefficients are presented in Table A.1). Linear associations between the MDI and measures of AL were variable. A consistent linear relation emerged only for actigraph AL in the test situation. This linear effect replicated across age and twin group (Tables A.6, A.8). Other significant linear relations did not replicate across age.

A more robust pattern emerged for quadratic effects, but only for observed AL. The significant R^2 for observed AL in the test and play situations at both ages indicates that adding the quadratic component to the models significantly increased variance explained in MDI. By contrast, when AL was assessed by actigraphs a curvilinear effect emerged only in the play situation at age 3 and explained just 1% of the variance in the MDI. These quadratic results for observed AL replicated across randomized twin groups (Tables A.6, A.8). The quadratic effect of actigraph AL in the play situation at age 3 approached significance ($p < .09$) in each twin group. For observed AL in the test situation, the model explained 8% of the variation in MDI scores at age 2, and 11.9% at age 3. Of this total variation, the quadratic component explained 75% (.06/.08) at age 2, and 92% (.11/.119) at age 3. Observed AL in the play situation explained 4% of the total variance in mental development at age 2, and 3.5% at age 3. Once again, most of this was due to the quadratic component (Age 2: [.04/.04=100%]; Age 3: [.03/.035=86%]). Figure 1 presents the fit plots showing the curvilinear relations between observed AL and MDI in the test and play situations at both ages. The 95% confidence intervals for the regression coefficients predicting MDI from observed AL in the test and play situations were overlapping, indicating that the curvilinear associations did not significantly differ across situation and that the link between AL and the MDI arises even when not assessed concurrently in the same situation. There was no curvilinear relation between parent-reported AL and the MDI.

Post-Hoc Longitudinal Analyses

To explore potential causality, we fit linear and quadratic models predicting MDI at age 3 from AL at age 2, controlling for MDI at age 2 (Table 4). Only AL variables with a linear or quadratic effect that replicated across age were included in these analyses (i.e., observed AL in test and play, actigraph AL in test). There were significant linear and curvilinear effects

between observed AL in the test situation at age 2 and the MDI at age 3; however, the effects replicated in only one randomized twin group, indicating that these cross-age effects are less robust (Table A.10). No significant longitudinal effects emerged for observed AL in the play situation, or for actigraph AL in the test situation, after controlling for age 2 MDI scores.

Discussion

This study investigated the possibility that AL may play a functional role in cognitive development (i.e., an “optimal” AL) by exploring both linear and curvilinear associations between the two domains in early childhood. Our use of a nonclinical sample allowed us to capture the full distribution of AL rather than simply the upper extreme. Thus we were able to examine both low and high levels of activity and their potential relation to mental development. A curvilinear relation between AL and mental development emerged, but only for observed AL. Nonetheless, these findings for AL observed within the lab replicated across situations (test and play), age (2 and 3 years), and samples (Random 1 and Random 2). Thus, there is strong support for the notion of an “optimal” AL in early child cognitive development when AL is assessed by knowledgeable observers within structured situations.

The curvilinear relation between AL and mental development is consistent with the Optimal Stimulation Model as well as with past research in clinical samples highlighting the utility of AL, even if not explicitly in the context of optimal stimulation (e.g., Rapport et al., 2009; Sarver, Rapport, Kofler, Raiker, & Friedman, 2015). These findings suggest that AL may support a cognitive function, or at least that at some moderate level is benign, but that there is a tipping point where too much or too little AL is deleterious. The negative implications of high AL on developmental outcomes are well-documented in the literature (Fagot & O’Brien, 1994; Rothbart et al., 1994), however, the effects of low AL have been given much less attention. To our knowledge, there are no prior studies looking at possible curvilinear relations between AL and developmental outcomes, but research with a related construct—ADHD behavior problems in a normative sample—found no curvilinear associations across multiple facets of development in adolescence (Greven et al., 2015). Rather, individuals with “low ADHD traits” had the most positive developmental outcomes. There are several possible reasons for the inconsistencies across studies. First, Greven et al. examined ADHD traits rather than AL. Even though the SWAN Rating Scale (Hay, Bennett, Levy, Sergeant, & Swanson, 2007) used to assess ADHD attempts to capture the full distribution, what is rated as “low” on activity-related items is the ability to modulate AL and not low levels of activity per se. Low activity is conceptualized on this measure, a priori, as a strength. Moreover, Greven et al. used the total ADHD score which combines hyperactivity, impulsivity and inattention into a single variable. Second, the participants in the study were adolescents. It is possible that the curvilinear pattern found in our study is specific to early childhood, during the important transition from infancy to childhood. Indeed, using AL as a means of addressing under or over arousal may decrease with age as other more internal ways (e.g., daydreaming) of addressing arousal can be utilized (Zentall & Zentall, 1983). Future research should investigate the possible curvilinear associations with AL at different developmental periods. Third, ADHD was measured via self-report and required adolescents to rate their behavior in the context of others (i.e., on a 7-point scale ranging from “far below average” to “far above average”). This degree of self-reflection may be difficult to do if one

does not have a firm understanding of what is normative. In contrast, our trained testers had extensive experience with a wide-range of children and were more able to make relative evaluations.

Despite replicating across situation, age and sample when AL was assessed by observation in the lab, a consistent curvilinear relation was not found between parent-rated AL or mechanically-assessed AL and mental development. To some extent, the discrepant results between observed and parent report is not surprising given that modest correlations between the two are reported in the literature (Saudino, Wertz, Gagne, & Chawla, 2004; Seifer, Sameroff, Barrett, & Krafchuk, 1994). Indeed, in the present sample, the correlations between parent ratings and observed AL, though significant, ranged from .10 to .19. These low correlations indicate that the two measures are tapping different aspects of AL. It is possible that the lab situation, particularly the test situation, elicits behaviors that are not typically seen by parents in day-to-day interactions. AL in the lab indexes motor movement and energy, but it may also be influenced by the novelty of the situation and increased attentional demands not often required at home. Similarly, assessing the child's reactions to mildly stressful lab situations may enrich observations of behaviors and make individual differences in AL more salient. Additionally, as with self-ratings of ADHD traits in adolescence, it is possible that parents have less relativity when rating their children's behaviors. Regardless of the source of the discrepancies between the two types of measures, our results highlight the importance of a multi-methodological approach in the investigation of childhood AL. Findings with one measure may not map onto another. While not possible in the current study, it would be interesting to explore the relations between teacher ratings of AL and developmental outcomes. Teachers are more objective raters of AL (Saudino, Ronald, & Plomin, 2005) and see this behavior within an academic setting, which might be closer to the types of behaviors we observed in the lab.

Although there were hints of a quadratic effect for actigraph AL in the play situation at age 3, when AL is mechanically-assessed the overall pattern of association with the MDI is linear, but only within the test situation. Because observers and actigraphs assessed AL within the same situations, simple contextual differences cannot explain the different patterns of association with the MDI. In contrast to parent ratings, the two measures were highly correlated. Nonetheless, correlations in the range of .56–.70 indicate that there is substantial variance unique to each measure. A plausible explanation for the differential effects lies in the fact that the actigraph provides only a *quantitative* measure of AL (i.e., movement frequency and intensity), whereas observer ratings of AL on the IBR also tap *qualitative* aspects of AL. That is, the observer does not simply rate the frequency of movement, but considers the level of activity within the circumstances of the task at hand and children are rated in relation to children who are the same age. As such, the same quantitative amount of activity observed within the play and test situations would be scored differently (e.g., a higher level of motoric activity would be more appropriate within the play situation). In addition, actigraphs provide an absolute measure of activity whereas observer ratings are relative. Although 2- and 3-year-olds differ in their absolute counts of motor activity (Saudino, 2012), because observers rate children relative to their same-aged peers, an “average” child would receive the same score at each age. Thus, observer ratings give some flavor to the measure of AL, which is lacking when actigraphs are employed. We

would suggest that “optimal activity” is not simply a matter of how active a child is, but rather the degree of activity within the boundaries of what would be appropriate to the situation and age.

Our somewhat inconsistent linear findings for AL mirror the inconsistent literature more broadly. It could be argued that previous studies captured different parts of the AL distribution and hence yielded different outcomes; however, we sampled the full distribution and found inconsistencies across our samples so this explanation does not apply. This points to the importance of large samples with increased power, replication and a multi-method approach. If we had examined the results from only one method, age, situation, or subsample, we would have come to very different conclusions about linear associations between AL and mental development. On the basis of our results, a cautious interpretation is that there is little evidence for a linear relation between the two in early childhood, with the exception of actigraph AL in the test situation. Given the modest contribution of the linear component, it is likely that previous studies were underpowered to reliably detect effects.

These findings have implications for how AL is conceptualized in early childhood. Across the transition from infancy to toddlerhood, high AL begins to be associated with more negative behavioral outcomes (e.g., Matheny, 1989) and as a result, is likely to be discouraged in many young children. However, the curvilinear pattern observed within our samples demonstrates that low AL within the context of the situation may also have negative developmental consequences. In early childhood, children with low AL may seem less problematic and troublesome to parents and teachers, but the current results suggest that they may be at risk for non-optimal developmental outcomes. Overall, moderate AL is associated with better mental development. Thus, all AL should not be viewed through a pejorative lens. The importance of acknowledging a normative and potentially developmentally-appropriate function of AL is highlighted not only by this study, but other research as well. For instance, Eaton et al. (2001) noted that the developmental peak in AL between 7 to 9 years of age corresponds to the peak in ADHD diagnoses, and suggested that this pattern hints that ADHD symptoms may be more normative than pathological. Additionally, increasing recess breaks in elementary school is related to less fidgeting and more on-task behavior in the classroom (e.g., Pellegrini, Huberty, & Jones, 1995). This is said to be because the higher AL that occurs as a result of more recess, indirectly or directly facilitates school performance (Eaton et al., 2001; Pellegrini & Bjorklund, 1997). Our results taken from AL observed during the play situation suggest that this would likely be the case only at moderate, normative amounts of AL.

The limitations of this study should be mentioned. First, the samples consisted of twins. It is possible that twins are not fully representative of the general population. For example, very young twins sometimes have lower cognitive abilities than singletons, but by preschool these differences have often righted themselves (Plomin et al., 2013). We restricted our sample to children who were not at risk due to low gestational age or birthweight, but we note that the mean scores for the MDI at age 2 were slightly lower than the population average of 100. However, this was not the case by age 3 and we get a similar pattern of results across age. Even if there were enduring mean differences between twin and singleton mental development scores in early childhood, the factors that influence mean differences can differ

from those that influence individual differences (Plomin et al., 2013). Hence, mean differences would not be expected to affect the pattern of associations between AL and MDI in twins and nontwins. Despite this, using a twin sample affords the unique opportunity to replicate results in a nearly perfectly matched sample. Nonetheless, replication in non-twins is encouraged. Second, as is usually the case with developmental research, the sample was predominantly white, middle-class, and well-educated. Replication with a more diverse sample is needed.

In conclusion, these results suggest an optimum level of AL with respect to mental development in early childhood. Contrary to the general assumption that high AL is bad, both high *and* low AL are associated with lower mental development, and children with moderate levels of activity as assessed by observers have the most optimal mental functioning. By examining AL across the full distribution and by applying a nonlinear statistical approach we are able to get a more complete understanding of children who are at developmental risk. Future researchers, clinicians, and educators should continue to focus on the entire continuum of AL in children, and consider the informative and meaningful heterogeneity within the entire population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The Boston University Twin Project (BUTP) is supported by grants MH062375 and HD068435 to Dr. Saudino. The twins' and families' participation is gratefully acknowledged.

References

- Bayley N (1969). Manual for the bayley scales of infant development New York: The Psychological Corporation.
- Bayley N (1993). Bayley scales of infant development: manual New York: The Psychological Corporation.
- Becker SP (2013). Topical review: Sluggish cognitive tempo: Research findings and relevance for pediatric psychology. *Journal of Pediatric Psychology*, 38(10), 1051–1057. [PubMed: 23933842]
- Becker SP, Ciesielski HA, Rood JE, Froehlich TE, Garner AA, Tamm L, & Epstein JN (2016). Uncovering a clinical portrait of sluggish cognitive tempo within an evaluation for attention-deficit/hyperactivity disorder: A case study. *Clinical Child Psychology and Psychiatry*, 21(1), 81–94. [PubMed: 25326531]
- Bjorklund DF (1997). The role immaturity in human development. *Psychological Bulletin*, 122(2), 153–169. [PubMed: 9283298]
- Brouwer SI, van Beijsterveldt TCEM, Bartels M, Hudziak JJ, & Boomsma DI (2006). Influences on Achieving Motor Milestones: A Twin-Singleton Study. *Twin Research and Human Genetics*, 9(3), 424–430. [PubMed: 16790152]
- Deater-Deckard K, Petrill SA, & Thompson LA (2007). Anger/frustration, task persistence, and conduct problems in childhood: a behavioral genetic analysis. *Journal of Child Psychology and Psychiatry*, 48(1), 80–87. [PubMed: 17244273]
- Eaton WO, McKeen NA, & Campbell DW (2001). The Waxing and Waning of Movement: Implications for Psychological Development. *Developmental Review*, 21(2), 205–223.

- Fagen JW, Singer JM, Ohr PS, & Fleckenstein LK (1987). Infant temperament and performance on the Bayley Scales of Infant Development at 4, 8, and 12 months of age. *Infant Behavior & Development*, 10(4), 505–512.
- Fagot BI, & O'Brien M (1994). Activity level in young children: Cross-age stability, situational influences, correlates with temperament, and the perception of problem behaviors. *Merrill-Palmer Quarterly*, 40(3), 378–398.
- Fair DA, Bathula D, Nikolas MA, & Nigg JT (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 109(17), 6769–6774.
- Goldsmith HH, Reilly J, Lemery KS, Longley S, & Prescott A (1995). *Preschool Laboratory Temperament Assessment Battery* Unpublished instrument, University of Wisconsin.
- Goldsmith HH (1996). Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Development*, 67(1), 218–235. [PubMed: 8605830]
- Graham JW (2012). Missing data: Analysis and design. In *Statistics for Social and Behavioral Sciences* (pp. 47–69). New York, NY: Springer Science & Business Media.
- Greven CU, Merwood A, Meer JMJ, Haworth CMA, Rommelse N, & Buitelaar JK (2015). The opposite end of the attention deficit hyperactivity disorder (adhd) continuum: Genetic and environmental aetiologies of extremely low adhd traits. *Journal of Child Psychology and Psychiatry*
- Halverson CF, & Waldrop MF (1976). Relations between preschool activity and aspects of intellectual and social behavior at age 7 1/2. *Developmental Psychology*, 12(2), 107–112.
- Hay DA, Bennett KS, Levy F, Sergeant J, & Swanson J (2007). A Twin Study of Attention-Deficit/Hyperactivity Disorder Dimensions Rated by the Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior. *Biological Psychiatry*, 61(5), 700–705. [PubMed: 16962074]
- Loo C, & Wenar C (1971). Activity level and motor inhibition: Their relationship to intelligence-test performance in normal children. *Child Development*, 42(3), 967–971. [PubMed: 5121107]
- Maccoby EE, Dowley EM, Hagen JW, & Degerman R (1965). Activity level and intellectual functioning in normal preschool children. *Child Development*, 36(3), 761–770.
- Martin RP, Olejnik S, & Gaddis L (1994). Is temperament an important contributor to schooling outcomes in elementary school? Modeling effects of temperament and scholastic ability on academic achievement. In Carey WB, McDevitt SC, Carey WB, & McDevitt SC (Eds.), *Prevention and early intervention: Individual differences as risk factors for the mental health of children: A festschrift for Stella Chess and Alexander Thomas* (pp. 59–68). Philadelphia, PA, US: Brunner/Mazel.
- Matheny AP (1980). Bayley's Infant Behavior Record: Behavioral components and twin analyses. *Child Development*, 51(4), 1157–1167. [PubMed: 7193557]
- Matheny AP (1989). Temperament and cognition: Relations between temperament and mental test scores. In Kohnstamm GA, Bates JE, Rothbart MK, Kohnstamm GA, Bates JE, & Rothbart MK (Eds.), *Temperament in childhood* (pp. 263–282). Oxford, England: John Wiley & Sons.
- Nettle D (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, 61(6), 622–631. [PubMed: 16953749]
- Pellegrini AD, & Bjorklund DF (1997). The role of recess in children's cognitive performance. *Educational Psychologist*, 32(1), 35–40.
- Pellegrini AD, Huberty PD, & Jones I (1995). The Effects of Recess Timing on Children's Playground and Classroom Behaviors. *American Educational Research Journal*, 32(4), 845–864.
- Piaget J, & Cook M (1952). *The origins of intelligence in children* New York, NY, US: W W Norton & Co.
- Plomin R, DeFries JC, Knopik VS, & Neiderheiser J (2013). *Behavioral genetics*, 6th edition. New York, NY: Worth.
- Rapport MD, Bolden J, Kofler MJ, Sarver DE, Raiker JS, & Alderson RM (2009). Hyperactivity in boys with attention-deficit/hyperactivity disorder (ADHD): A ubiquitous core symptom or manifestation of working memory deficits? *Journal of Abnormal Child Psychology*, 37(4), 521–534. [PubMed: 19083090]

- Rothbart MK, Derryberry D, & Hershey K (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In Molfese VJ, Molfese DL, Molfese VJ, & Molfese DL (Eds.), *Temperament and personality development across the life span* (pp. 85–119). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Rothbart MK, Derryberry D, & Posner MI (1994). A psychobiological approach to the development of temperament. In Bates JE, Wachs TD, Bates JE, & T. Wachs D (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 83–116). Washington, DC, US: American Psychological Association.
- Sarver DE, Rapport MD, Kofler MJ, Raiker JS, & Friedman LM (2015). Hyperactivity in Attention-Deficit/Hyperactivity Disorder (ADHD): Impairing Deficit or Compensatory Behavior? *Journal of Abnormal Child Psychology*, 43(7), 1219–1232. [PubMed: 25863472]
- Saudino KJ (2012). Sources of continuity and change in activity level in early childhood. *Child Development*, 83(1), 266–281. [PubMed: 22103336]
- Saudino KJ, & Cherny SS (2001). Sources of continuity and change in observed temperament. In Emde RN, Hewitt JK, Emde RN, & Hewitt JK (Eds.), *Infancy to early childhood: Genetic and environmental influences on developmental change* (pp. 89–110). New York, NY, US: Oxford University Press.
- Saudino KJ, Ronald A, & Plomin R (2005). The Etiology of Behavior Problems in 7-Year-Old Twins: Substantial Genetic Influence and Negligible Shared Environmental Influence for Parent Ratings and Ratings by Same and Different Teachers. *Journal of Abnormal Child Psychology*, 33(1), 113–130. [PubMed: 15759595]
- Saudino KJ, Wertz AE, Gagne JR, & Chawla S (2004). Night and Day: Are Siblings as Different in Temperament as Parents Say They Are? *Journal of Personality and Social Psychology*, 87(5), 698–706. [PubMed: 15535780]
- Saudino KJ, & Zapfe JA (2008). Genetic influences on activity level in early childhood: Do situations matter? *Child Development*, 79(4), 930–943. [PubMed: 18717899]
- Seifer R, Sameroff AJ, Barrett LC, & Krafchuk E (1994). Infant temperament measured by multiple observations and mother report. *Child Development*, 65(5), 1478–1490. [PubMed: 7982363]
- Vazou S, & Smiley-Oyen A (2014). Moving and Academic Learning Are Not Antagonists: Acute Effects on Executive Function and Enjoyment. *Journal of Sport and Exercise Psychology*, 36(5), 474–485. [PubMed: 25356611]
- Zentall SS, & Zentall TR (1983). Optimal stimulation: A model of disordered activity and performance in normal and deviant children. *Psychological Bulletin*, 94(3), 446–471. [PubMed: 6657825]

Key Points

- Research on activity level (AL) has typically focused on the upper end and ignores the possibility that low AL may also be detrimental.
- We are the first to demonstrate a curvilinear relation between AL and mental development in a normative sample. Results replicated across matched samples, contexts and age.
- In early childhood, both high and low AL is detrimental, and moderate AL may serve a function.
- It is important to consider the meaningful heterogeneity across the entire distribution in both research and clinical work.

Figure 1.
Fit Plots of Relation Between AL and MDI Across Age and Situation

Table 1.

Descriptive Statistics

	Age 2 <i>M (SD)</i> <i>n</i> =622	Age 3 <i>M (SD)</i> <i>n</i> =600
MDI	90.76 (14.60) <i>n</i> =622	98.60 (12.14) <i>n</i> =600
Observed-Test AL	0.01 (0.89) <i>n</i> =626	0.00 (0.91) <i>n</i> =607
Observed-Play AL	0.00 (0.90) <i>n</i> =626	0.00 (0.90) <i>n</i> =607
Actigraph-Test AL	571.68 (234.31) <i>n</i> =625	493.47 (209.94) <i>n</i> =605
Actigraph-Play AL	778.57 (344.89) <i>n</i> =624	723.28 (270.91) <i>n</i> =605
Actigraph-Home AL	451.85 (102.21) <i>n</i> =623	503.41 (107.32) <i>n</i> =605
Parent-rated AL	41.71 (8.68) <i>n</i> =617	39.54 (9.04) <i>n</i> =592

Table 2.

Concurrent intercorrelations between MDI and AL variables collapsed across groups

	MDI	Observed-Test AL	Observed-Play AL	Actigraph-Test AL	Actigraph-Play AL	Actigraph-Home AL	Parent- rated AL
MDI	-	-0.18 **	-0.11 **	-0.29 **	-0.16 **	-0.08 *	-0.10 *
Observed-Test AL	-0.11 **	-	0.64 **	0.63 **	0.46 **	0.15 **	0.16 **
Observed-Play AL	-0.09 *	0.67 **	-	0.46 **	0.68 **	0.17 **	0.19 **
Actigraph-Test AL	-0.22 **	0.56 **	0.45 **	-	0.64 **	0.34 **	0.16 **
Actigraph-Play AL	-0.05	0.50 **	0.70 **	0.59 **	-	0.38 **	0.13 **
Actigraph-Home AL	-0.04	0.16 **	0.15 **	0.31 **	0.30 **	-	0.25 **
Parent-rated AL	-0.19 **	0.19 **	0.09 *	0.21 **	0.15 **	0.19 **	-

Note. Values above the diagonal are for AL variables at Age 2, and values below the diagonal are for AL variables at Age 3;

* $p < 0.05$;

** $p < 0.01$.

Table 3.

Concurrent unstandardized linear and quadratic model results predicting MDI from AL

Age 2 Activity Measure	Linear <i>B</i> [95% CI]	Quadratic <i>B</i> [95% CI]	Linear <i>R</i> ²	<i>R</i> ²	<i>F</i> , <i>df</i>
Observed-Test	-2.64 [-4.02; -1.25]	-3.39 [-4.56; -2.22]	0.02 **	0.06 **	23.42, (2, 313)
Observed-Play	-1.71 [-3.17; -0.25]	-2.96 [-4.12; -1.80]	0.00	0.04 **	13.02, (2, 313)
Actigraph-Test	-0.75 [-1.01; -0.49]	-0.01 [-0.05; 0.03]	0.06 **	0.00	16.37, (2, 313)
Actigraph-Play	-0.24 [-0.45; -0.04]	-0.01 [-0.04; 0.01]	0.01 *	0.00	3.96, (2, 313)
Actigraph-Home	-0.36 [-0.95; 0.23]	0.02 [-0.17; 0.21]	0.00	0.00	0.76, (2, 312)
Parent-rated	-0.09 [-0.25; 0.07]	-0.01 [-0.02; 0.00]	0.00	0.00	1.97, (2, 310)
Age 3 Activity Measure	Linear <i>B</i> [95% CI]	Quadratic <i>B</i> [95% CI]	Linear <i>R</i> ²	<i>R</i> ²	<i>F</i> , <i>df</i>
Observed-Test	-1.07 [-2.26; 0.13]	-3.25 [-4.18; -2.32]	0.01	0.11 **	28.52, (2, 302)
Observed-Play	-1.41 [-2.70; -0.13]	-1.56 [-2.48; -0.64]	0.01	0.03 **	8.02, (2, 302)
Actigraph-Test	-0.53 [-0.75; -0.31]	-0.02 [-0.05; 0.01]	0.04 **	0.00	12.53, (2, 302)
Actigraph-Play	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.00	0.01 **	6.22, (2, 302)
Actigraph-Home	0.00 [-0.01; 0.01]	0.00 [0.00; 0.00]	0.00	0.00	0.47, (2, 301)
Parent-rated	-0.22 [-0.36; -0.09]	0.00 [-0.01; 0.01]	0.03 **	0.00	6.88, (2, 294)

Note.

*
p < 0.05;**
p < 0.01.The *F* statistic and *df* are presented for the overall quadratic model.

Table 4.

Longitudinal unstandardized linear and quadratic model results predicting MDI at age 3 from AL at age 2

Age 2 Activity Measure	MDI Linear <i>B</i> [95% CI]	AL Linear <i>B</i> [95% CI]	AL Quadratic <i>B</i> [95% CI]	Linear MDI <i>R</i> ²	AL Linear <i>R</i> ²	AL Quadratic <i>R</i> ²	<i>F</i> , <i>df</i>
Observed-Test	.51 [.44; .57]	-1.41 [-2.42; -.40]	-1.16 [-2.10; -.27]	.39 **	0.01 *	0.01 *	109.98, (3, 302)
Observed-Play	.53 [.46; .59]	-.54 [-1.55; .47]	-.41 [-1.29; .47]	.39 **	0.001	0.001	98.01, (3, 302)
Actigraph-Test	.53 [.46; .60]	-.09 [-.29; .11]	-.01 [-.03; .02]	.39 **	0.001	0.0002	93.24, (3, 302)

Note.

*
 $p < 0.05$;**
 $p < 0.01$.