Attention-deficit/hyperactivity disorder and sluggish cognitive tempo throughout childhood: temporal invariance and stability from preschool through ninth grade

Daniel R. Leopold,1,2 Micaela E. Christopher,1,2 G. Leonard Burns,3 Stephen P. Becker,4 Richard K. Olson,1,2 and Erik G. Willcutt1,2

1Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA; 2Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA; 3Department of Psychology, Washington State University, Pullman, WA, USA; 4Division of Behavior Medicine and Clinical Psychology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA

Background: Although multiple cross-sectional studies have shown symptoms of sluggish cognitive tempo (SCT) and attention-deficit/hyperactivity disorder (ADHD) to be statistically distinct, studies have yet to examine the temporal stability and measurement invariance of SCT in a longitudinal sample. To date, only six studies have assessed SCT longitudinally, with the longest study examining SCT over a 2-year period. The overall goals of this study were to assess the 10-year longitudinal stability and interfactor relationships of ADHD and SCT symptoms among a community sample of children.

Methods: Confirmatory factor analysis was used to assess the temporal invariance of ADHD and SCT symptoms in a large population-based longitudinal sample (International Longitudinal Twin Study of Early Reading Development) that included children assessed at preschool and after kindergarten, first, second, fourth, and ninth grades (n = 489). Latent autoregressive models were then estimated to assess the stability of these constructs. Results: Results demonstrated invariance of item loadings and intercepts from preschool through ninth grades, as well as invariance of interfactor correlations. Results further indicated that both ADHD and SCT are highly stable across these years of development, that these symptom dimensions are related but also separable, and that hyperactivity/impulsivity and SCT are both more strongly correlated with inattention than with each other and show differential developmental trajectories. Specifically, even in the presence of latent simplex analyses providing support for the developmental stability of these dimensions, linear comparisons indicated that that mean levels of hyperactivity/impulsivity decreased with time, inattentive ratings were generally stable, and SCT tended to increase slightly across development.

Conclusions: This study adds to the current literature by being the first to systematically assess and demonstrate the temporal invariance and stability of ADHD and SCT across a span of 10 years. Keywords: ADHD; attention-deficit/hyperactivity disorder; sluggish cognitive tempo; sluggish tempo; inattention; hyperactivity; stability; temporal invariance; measurement invariance.

Introduction

The notion that a cluster of symptoms related to, but distinct from, inattention (IN) might reliably distinguish between individuals with and without hyperactivity/impulsivity (HI) symptoms has been around for almost three decades (Nepper & Lahey, 1986). This symptom cluster, composed of items such as sluggishness, being ‘lost in a fog’, and daydreaming, has been termed sluggish cognitive tempo (SCT). SCT was initially identified as a set of symptoms separable from the IN symptoms of attention-deficit disorder from the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III; American Psychiatric Association [APA], 1980). Although SCT-related items were included in the DSM-IV (American Psychiatric Association, 1994) attention-deficit/hyperactivity disorder (ADHD) field trials (Frick et al., 1994), research on SCT largely remained dormant until the turn of the century, when more investigations turned to examining SCT in its own right (Becker, Marshall, & McBurnett, 2014). Thus, studies of SCT are increasingly moving away from examining its prevalence and correlates among ADHD-selected samples and instead exploring these two constructs side-by-side (Barkley, 2012, 2013; Bernad, Servera, Grases, Collado, & Burns, 2014).

Despite strong support for the validity of the ADHD diagnosis (Willcutt et al., 2012), SCT is in a comparatively nascent state. As the body of SCT literature continues to grow, it is imperative that we test whether SCT meets the criteria for the validation of a mental disorder (e.g., Cantwell, 1980). Although SCT has been described as a disorder and alternative names have been proposed (Barkley, 2014), these proposals run the risk of pathologizing nonpathological behaviors and jeopardizing public perception of psychology and psychiatry. It is therefore important to put the extant literature into perspective and systematically assess these criteria before recommending its diagnostic status and inclusion in...
clinical practice. As such, the current study aimed to evaluate fundamental nosological requirements, namely the temporal stability, measurement invariance, and developmental trajectory of SCT and its relationship with ADHD.

Validity of SCT in children and adolescents

Factor analyses and external correlates. Initial and ongoing factor analyses have repeatedly supported the separability of SCT and ADHD, as well as SCT’s comparatively stronger relationship with symptoms of IN than with symptoms of HI (Barkley, 2013; Leopold, Bryan, Pennington, & Willcutt, 2014; Penny, Waschbusch, Klein, Corkum, & Eskes, 2009; Willcutt et al., 2012). Numerous studies have subsequently found that symptoms of SCT are associated with aspects of functional impairment and elevated internalizing symptoms, even after controlling for symptoms of ADHD, related psychopathologies, and demographic characteristics (Becker, Langberg, Luebbe, Dvorsky, & Flannery, 2013; Lee, Burns, Snell, & McBurnett, 2014; Willcutt et al., 2014). Of note, a smaller body of literature has also begun to explore the SCT construct among adults (Barkley, 2012; Becker et al., 2013; Leopold et al., 2014), as well as the presence of rater effects and the importance of collecting ratings from multiple sources (Burns, Servera, Bernad, Carrillo, & Cardo, 2013; Lee et al., 2014; Leopold et al., 2014).

Temporal stability. Although the categorical ADHD subtype classifications described in DSM-IV are developmentally unstable (Willcutt et al., 2012), the DSM-IV symptom dimensions and overall ADHD diagnosis demonstrate high stability throughout the life span, with diagnostic symptoms and impairment extending into adulthood for 50–80% of cases (Barkley, 2013; Wilens, Faraone, & Biederman, 2004). Test–retest reliability of the IN and HI dimensions ranges from \( r = .78 \) to \( r = .82 \) for \(< 1\) year and \( r = .64 \) for these dimensions for intervals of 1–5 years (Willcutt et al., 2012). The ADHD dimensions also show differential longitudinal courses, with IN remaining relatively stable, whereas HI shows a significant age-related decline, particularly during the early school years (e.g., Lahey & Willcutt, 2010).

In comparison with the large body of literature on ADHD across the life span, only six studies to date have investigated the developmental stability of SCT. Over a period of 4–6 weeks, test–retest reliability estimates of parents’ ratings range from \( r = .73 \) to \( r = .80 \) using both latent factors (Burns et al., 2013) and composites (Lee et al., 2014). Using teacher ratings of 176 first to sixth graders, Becker found that Penny et al.’s (2009) SCT scale \( \bar{z}_{\text{initial}} = .92; \bar{z}_{\text{follow-up}} = .95 \) had a 6-month test–retest correlation of \( r = .76 \) (unpublished data from Becker, 2014). These results support Penny and colleagues’ initial 3-month test–retest reliability of \( r = .87 \) for parents’ ratings. Bernad et al. (2014) found that three commonly used SCT items – seems drowsy, thinking is slow, and slow moving – demonstrated both convergent and discriminant validity as well as temporal stability. Specifically, in a sample of primary and secondary teachers’ reports of Spanish children’s behavior, Bernad et al. found that the 1-year test–retest correlations were \( r = .74 \) and \( .61 \), respectively. A recent study followed this same sample for an additional year and, using different teachers across time points, reported 2-year latent test–retest correlations of \( r = .46 \) and \( .42 \) for primary and secondary teachers, respectively; among mothers and fathers, the 2-year latent test–retest correlations (using the three aforementioned SCT items in addition to loses train of thought and easily confused) were \( r = .60 \) and \( .65 \), respectively (Bernad, Servera, Becker, & Burns, 2015). Thus, across parents, caregivers, and primary and secondary teachers, SCT appears to be a highly stable trait over these 6-week to 2-year timespans.

Measurement invariance. To ensure that a construct can be reliably measured at different time points, it is necessary to demonstrate that the construct’s items and variances are stable (i.e., invariant) across time. Confirmatory factor analyses of ADHD have previously demonstrated this important quality of temporal and measurement invariance across settings and raters (e.g., Burns, Servera, Bernad, Carrillo, & Geiser, 2014), but the temporal measurement invariance of SCT has yet to be investigated.

The current study

The current study sought to address this lack of longitudinal research on SCT by using parent ratings of IN, HI, and SCT collected six times between preschool and ninth grade among a representative, community sample of 489 children. Furthermore, the current study uses structural equation modeling (SEM) to explore the temporal invariance and longitudinal stability of these dimensions by using latent variables that minimize occasion-specific measurement error. The study had three main objectives:

1. To clarify the relations between IN, HI, and SCT, initial CFAs were fit to parent ratings at each of the six time points prior to testing the temporal invariance of these dimensions from preschool through the end of ninth grade. Based on previous findings, we hypothesized that correlated dimensions of IN, HI, and SCT symptoms would fit the data well at all ages, providing support for the measurement invariance of SCT across a 10-year span.

2. Cross-sectional interfactor and longitudinal intrafactor correlations were computed to explore the relations between SCT, IN, and HI, as well as their stability across development. If parent ratings show stability both between and within these
dimensions over time, the findings will provide key support for the internal and discriminant validity of ADHD and SCT during childhood and adolescence.

3. Finally, latent simplex (i.e., autoregressive) models of the IN, HI, and SCT dimensions were tested to explore the stability of these traits across all time points by only modeling the shared, reliable variance among symptoms. High path coefficients between time points would indicate that the overall rank order of individuals in this community sample remained consistent.

**Methods**

**Participants**

The participants in the present study were 489 individuals from 224 monozygotic (MZ; i.e., identical) and 265 dizygotic (DZ; i.e., fraternal) twin pairs first assessed in the year prior to entering kindergarten ($N = 482$, $M_{age} = 4.89$ years, $SD_{age} = 0.41$ years; 49.7% male). For all analyses presented, one randomly selected member from each pair was used to eliminate issues related to nonindependence. All participants were part of the Colorado component of the International Longitudinal Twin Study of Early Reading Development (ILTSERD; e.g., Christopher et al., 2013). All twin pairs were recruited from the Colorado Twin Registry based on birth records and had English as their first language. After the initial preschool wave, the participants were assessed in the summers following kindergarten ($N = 453$, $M_{age} = 6.3$ years, $SD_{age} = .31$ years), first grade ($N = 442$, $M_{age} = 7.4$ years, $SD_{age} = .32$ years), second grade ($N = 451$, $M_{age} = 8.5$ years, $SD_{age} = .31$ years), fourth grade ($N = 445$, $M_{age} = 10.5$ years, $SD_{age} = .32$ years), and ninth grade ($N = 389$, $M_{age} = 15.4$ years, $SD_{age} = .30$ years). Attrition from preschool through the end of fourth grade was minimal (92–94% retention). The current rate of retention is similar for the ninth-grade assessment (92%), but the sample size for this wave is smaller because one of the five cohorts had not yet completed the ninth-grade assessment. Importantly, this last wave is smaller because one of the five cohorts had not yet completed the ninth-grade assessment. First, the twins completed a battery of measures related to reading development in individual testing sessions, whereas one parent or caregiver completed a battery of questionnaires that included the measures described in this report.

**Procedures**

All study procedures were fully approved by the Institutional Review Boards of the University of Colorado Boulder. All participants and parents provided informed consent or assent prior to enrollment and at each follow-up assessment. Overall testing procedures for the ILTSERD are described in detail in previous papers (e.g., Christopher et al., 2013). Briefly, the twins completed a battery of measures related to reading development in individual testing sessions, whereas one parent or caregiver completed a battery of questionnaires that included the measures described in this report.

**Measures**

**DSM-IV ADHD symptoms.** The Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998) was used to obtain parent ratings of ADHD symptoms. Nearly all ratings were completed by the mother at all time points (91–95%). The child’s parent is asked to indicate how often in the last 6 months each of the 18 DSM-IV ADHD symptoms is true on a 4-point Likert scale (0 = never or rarely, 1 = sometimes, 2 = often, and 3 = very often).

**Sluggish cognitive tempo.** Seven potential SCT items (sluggish, slow to respond, lethargic; seems not to hear, needs things repeated; seems to ‘be in a fog’; is drowsy or sleepy; easily confused; daydreams, stares into space; and absent-minded, forgets things easily) were used based on previous studies and theoretical models of SCT (e.g., Penny et al., 2009). Each item was added to the ADHD rating scale and administered in the same four-point scale format.

**Analytic strategy**

Structural and measurement model analyses were conducted using Mplus (version 7.0; Muthén & Muthén, 1998–2012). Items were treated as ordered categorical manifest variables using the robust weighted least-squares estimator (WLSMV). Overall model fit was determined with the robust comparative fit index (CFI; study criteria ≥.90, with approximately .95 being ideal), the Tucker-Lewis index (TLI; study criteria ≥.90), and the robust root-mean-square error of approximation (RMSEA; study criteria ≤.08).

**Criteria and procedure for invariance tests.** Due to the ability of the chi-square difference test to detect small discrepancies of no theoretical or practical consequence in large samples (Chen, Sousa, & West, 2005), changes in CFI, TLI, and RMSEA were used to assess the invariance of model constraints. Specifically, if the decrease in CFI was <.01 and the TLI and RMSEA showed little change (Chen, 2007; Little, 2013), the imposed constraints were assumed to be invariant by this procedure. To reduce model complexity while maintaining the full 10-year timespan, only the preschool, post-first-, post-fourth-, and post-ninth-grade time points were used for the invariance analyses.

**CFA of temporal invariance.** The baseline model contains 12 factors and 100 items (i.e., 25 items on three factors at each time point – 9 IN, 9 HI, and 7 SCT). We now outline the sequence of steps used to test the temporal invariance of IN, HI, and SCT at these four time points (see Chapter 5 of Little, 2013 for a similar model). To evaluate configural invariance with correlated errors for identical items, none of the model parameters are constrained to be equal across time points in the baseline model. Furthermore, because the residual variance of each item was expected to covary with that same item’s residual variance at the other time points (Little, 2013), correlated errors were estimated between identical items at the four time points (see Figure 1). To evaluate invariance of loadings and intercepts (i.e., comparing raw and latent means across time points), it is first necessary to demonstrate the invariance of loadings and intercepts (i.e., thresholds for categorical items) for the same items rated at different time points (Chen et al., 2005). If statistical equivalence of items’ loadings and thresholds is supported, it can be inferred that observed score differences between time points reflect true differences on the dimensions’ means and not the effect of the measurement occasion itself. Finally, to assess the stability and strength of relationships between the IN, HI, and SCT dimensions, the tenability of factor correlation invariance was assessed. To impose these constraints, phantom constructs were used to estimate the variance-covariance information as standard deviations and correlations. When the variances of constructs differ between measurement occasions or groups, phantom constructs conveniently model the strength of these associations on a
common metric rather than the noncomparable metric of the
covariance relationships (Little, 2013).

Bivariate and longitudinal simplex analyses. After
establishing the measurement invariance of these three con-
structs across the 10-year timespan, longitudinal intracon-
struct and cross-sectional interconstruct bivariate correlations
were calculated (see Table S1, available online). Finally,
because the measured bivariate correlations indicated that
correlations between adjoining time points were stronger than
correlations between distant time points (Table S1), latent
simplex (i.e., autoregressive) models were estimated to explore
the temporal stability of IN, HI, and SCT, free from measure-
ment error.

Missing data. As noted earlier, attrition from preschool
through post-fourth grade was minimal; roughly 9% of the
sample was lost to follow up. The post-ninth-grade sample,
however, had not been fully collected, so 80% of the original
sample was available at this time. Factor correlations among
this reduced post-ninth-grade sample are similar to the other
time points. Covariance coverage for preschool to post-ninth
grade ranged from .73 to .99 (M = .85, SD = .09). Analyses
with the WLSMV estimator use a pairwise approach to
missing data.

Results

Factor structure and internal consistency

Preliminary CFAs in which each hypothesized latent
construct predicted its respective indicators were
estimated using the factor variance method of iden-
tification. The three-factor models provided a rea-
sonable fit at each time point, providing support for
estimation of the full configural invariance model
(results available from first author upon request).

Individual IN, HI, and SCT items loaded strongly
on their corresponding latent trait at each of the time
points (mean standardized loadings = .79 for IN, .75
for HI, and .77 for SCT), and estimates of internal
consistency were high for composite measures of IN,
HI, and SCT (mean $\alpha = .90, .87$, and .81, respec-
tively; Table 1).

Table 1 Descriptive statistics and internal reliability of parent
ratings of IN, HI, and SCT

<table>
<thead>
<tr>
<th>Measure/wave</th>
<th>$M$</th>
<th>$SD$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preschool</td>
<td>.67</td>
<td>.50</td>
<td>.89</td>
</tr>
<tr>
<td>Post-K</td>
<td>.56</td>
<td>.48</td>
<td>.89</td>
</tr>
<tr>
<td>Post-1st</td>
<td>.57</td>
<td>.50</td>
<td>.90</td>
</tr>
<tr>
<td>Post-2nd</td>
<td>.59</td>
<td>.51</td>
<td>.90</td>
</tr>
<tr>
<td>Post-4th</td>
<td>.65</td>
<td>.56</td>
<td>.92</td>
</tr>
<tr>
<td>Post-9th</td>
<td>.63</td>
<td>.61</td>
<td>.93</td>
</tr>
<tr>
<td>Preschool</td>
<td>.78</td>
<td>.55</td>
<td>.87</td>
</tr>
<tr>
<td>Post-K</td>
<td>.55</td>
<td>.51</td>
<td>.88</td>
</tr>
<tr>
<td>Post-1st</td>
<td>.49</td>
<td>.49</td>
<td>.88</td>
</tr>
<tr>
<td>Post-2nd</td>
<td>.45</td>
<td>.46</td>
<td>.88</td>
</tr>
<tr>
<td>Post-4th</td>
<td>.40</td>
<td>.44</td>
<td>.87</td>
</tr>
<tr>
<td>Post-9th</td>
<td>.33</td>
<td>.43</td>
<td>.86</td>
</tr>
<tr>
<td>Preschool</td>
<td>.22</td>
<td>.29</td>
<td>.75</td>
</tr>
<tr>
<td>Post-K</td>
<td>.21</td>
<td>.32</td>
<td>.80</td>
</tr>
<tr>
<td>Post-1st</td>
<td>.20</td>
<td>.29</td>
<td>.76</td>
</tr>
<tr>
<td>Post-2nd</td>
<td>.24</td>
<td>.35</td>
<td>.82</td>
</tr>
<tr>
<td>Post-4th</td>
<td>.29</td>
<td>.42</td>
<td>.88</td>
</tr>
<tr>
<td>Post-9th</td>
<td>.35</td>
<td>.45</td>
<td>.86</td>
</tr>
</tbody>
</table>

IN, inattention; HI, hyperactivity/impulsivity; SCT, sluggish
cognitive tempo.

Invariance of the three-factor model from preschool
through post-ninth grade

Table 2 shows that the baseline configural invariance
model with correlated errors between like items
displayed an excellent fit to the present data
($CFI = .95; TLI = .95; RMSEA = .025 [90\% CI = .023,
.026]$). The model remained invariant with the load-
ing, threshold, and latent factor correlation con-
straints imposed (see Table 2); all changes in fit
indices were below the critical values suggested by
Chen (2007) and Little (2013). Furthermore, none of
these changes from the baseline to final model
approached the critical values of .01 (CFI), .01
(TLI), and .015 (RMSEA).

© 2016 Association for Child and Adolescent Mental Health.
Mean levels of HI symptoms declined significantly across development (Table 1), with medium to large effect sizes for the difference between parents ratings obtained in preschool and after kindergarten and ratings obtained after fourth and ninth grades ($d = .34–.94$). Mean levels of IN during this timespan generally remained stable (all $d < .17$), whereas mean levels of SCT demonstrated small but significant increases, particularly from second to fourth and fourth to ninth grades ($d s = .18–.31$).

Table 3 shows the latent variance–covariance and correlation matrixes from the strong invariant model. With the addition of constraints on the latent interfactor correlations across all time points, the standardized correlations between IN and HI, IN and SCT, and HI and SCT were .68, .83, and .53, respectively.

As noted by Brown (2015, p.116) and Little (2013), it is reasonable to question whether factors represent separate constructs as the interfactor correlation approaches 1.0. In particular, they cite $r = .85$ as an upper bound for the demonstration of discriminant validity. To test whether our latent correlation of .83 is significantly less than 1.0, we directly compared our model against one in which the correlation was set to 1.0. A highly significant difference test, $\chi^2 (1, N = 489) = 174.6, p < .001$ indicates that our original model with separable IN and SCT factors demonstrates superior fit. This latent correlation of .83 means that 31% of the true score variance of the IN and SCT factors is nonoverlapping. Thus, although IN and SCT are highly correlated, they still satisfy the minimum conditions for discriminant validity.

Table 2 Invariance of parent-rated IN, HI, and SCT at preschool, post-1st, post-4th, and post-9th grades

<table>
<thead>
<tr>
<th>Model tested</th>
<th>$df$</th>
<th>$\chi^2$</th>
<th>$p$</th>
<th>$\Delta df$</th>
<th>$\Delta \chi^2$</th>
<th>$p$</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA [90% CI]</th>
<th>MC</th>
<th>ΔCFI</th>
<th>ATLI</th>
<th>Pass?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>4950</td>
<td>36649.9</td>
<td>&lt;.001</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Measurement model estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: Configural invariance; correlated errors across like items</td>
<td>4634</td>
<td>6191.7</td>
<td>&lt;.001</td>
<td>–</td>
<td>–</td>
<td>.95</td>
<td>.95</td>
<td>.026</td>
<td>[.024, .028]</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>2: Loading and threshold invariance (strong invariance)</td>
<td>4829</td>
<td>6440.7</td>
<td>&lt;.001</td>
<td>–</td>
<td>–</td>
<td>.95</td>
<td>.95</td>
<td>.026</td>
<td>[.024, .028]</td>
<td>2 vs. 1</td>
<td>.002</td>
<td>.000</td>
<td>Yes</td>
</tr>
<tr>
<td>3: Latent factor correlation invariance</td>
<td>4838</td>
<td>6435.4</td>
<td>&lt;.001</td>
<td>9</td>
<td>.064</td>
<td>.95</td>
<td>.95</td>
<td>.026</td>
<td>[.024, .028]</td>
<td>3 vs. 2</td>
<td>.001</td>
<td>.000</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 3 Latent variance–covariance matrix and latent means from strong factorial invariance model

<table>
<thead>
<tr>
<th>Construct</th>
<th>IN T1</th>
<th>HI T1</th>
<th>SCT T1</th>
<th>IN T2</th>
<th>HI T2</th>
<th>SCT T2</th>
<th>IN T3</th>
<th>HI T3</th>
<th>SCT T3</th>
<th>IN T4</th>
<th>HI T4</th>
<th>SCT T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN T1</td>
<td>.81</td>
<td>.78</td>
<td>.82</td>
<td>.64</td>
<td>.56</td>
<td>.63</td>
<td>.60</td>
<td>.55</td>
<td>.55</td>
<td>.59</td>
<td>.49</td>
<td>.57</td>
</tr>
<tr>
<td>HI T1</td>
<td>.62</td>
<td>.79</td>
<td>.62</td>
<td>.51</td>
<td>.73</td>
<td>.47</td>
<td>.48</td>
<td>.67</td>
<td>.40</td>
<td>.47</td>
<td>.63</td>
<td>.42</td>
</tr>
<tr>
<td>SCT T1</td>
<td>.61</td>
<td>.45</td>
<td>.68</td>
<td>.66</td>
<td>.53</td>
<td>.78</td>
<td>.54</td>
<td>.44</td>
<td>.62</td>
<td>.44</td>
<td>.33</td>
<td>.60</td>
</tr>
<tr>
<td>IN T2</td>
<td>.55</td>
<td>.43</td>
<td>.51</td>
<td>.90</td>
<td>.75</td>
<td>.92</td>
<td>.80</td>
<td>.61</td>
<td>.75</td>
<td>.63</td>
<td>.49</td>
<td>.65</td>
</tr>
<tr>
<td>HI T2</td>
<td>.52</td>
<td>.67</td>
<td>.45</td>
<td>.73</td>
<td>1.06</td>
<td>.68</td>
<td>.62</td>
<td>.84</td>
<td>.53</td>
<td>.44</td>
<td>.66</td>
<td>.43</td>
</tr>
<tr>
<td>SCT T2</td>
<td>.53</td>
<td>.39</td>
<td>.60</td>
<td>.81</td>
<td>.65</td>
<td>.87</td>
<td>.78</td>
<td>.56</td>
<td>.84</td>
<td>.61</td>
<td>.42</td>
<td>.74</td>
</tr>
<tr>
<td>IN T3</td>
<td>.56</td>
<td>.44</td>
<td>.46</td>
<td>.78</td>
<td>.66</td>
<td>.75</td>
<td>1.08</td>
<td>.76</td>
<td>.92</td>
<td>.75</td>
<td>.55</td>
<td>.72</td>
</tr>
<tr>
<td>HI T3</td>
<td>.53</td>
<td>.64</td>
<td>.38</td>
<td>.62</td>
<td>.92</td>
<td>.56</td>
<td>.84</td>
<td>1.15</td>
<td>.67</td>
<td>.51</td>
<td>.72</td>
<td>.43</td>
</tr>
<tr>
<td>SCT T3</td>
<td>.55</td>
<td>.40</td>
<td>.57</td>
<td>.80</td>
<td>.62</td>
<td>.88</td>
<td>1.07</td>
<td>.80</td>
<td>1.27</td>
<td>.70</td>
<td>.46</td>
<td>.79</td>
</tr>
<tr>
<td>IN T4</td>
<td>.60</td>
<td>.48</td>
<td>.41</td>
<td>.67</td>
<td>.51</td>
<td>.64</td>
<td>.88</td>
<td>.62</td>
<td>.89</td>
<td>1.28</td>
<td>.68</td>
<td>.90</td>
</tr>
<tr>
<td>HI T4</td>
<td>.52</td>
<td>.65</td>
<td>.32</td>
<td>.54</td>
<td>.79</td>
<td>.45</td>
<td>.67</td>
<td>.90</td>
<td>.60</td>
<td>.90</td>
<td>1.35</td>
<td>.56</td>
</tr>
<tr>
<td>SCT T4</td>
<td>.60</td>
<td>.44</td>
<td>.58</td>
<td>.72</td>
<td>.51</td>
<td>.81</td>
<td>.88</td>
<td>.54</td>
<td>1.04</td>
<td>1.18</td>
<td>.76</td>
<td>1.37</td>
</tr>
<tr>
<td>Means:</td>
<td>.05</td>
<td>.05</td>
<td>.08</td>
<td>.17</td>
<td>.66</td>
<td>.07</td>
<td>.05</td>
<td>.95</td>
<td>.18</td>
<td>.26</td>
<td>1.51</td>
<td>.09</td>
</tr>
</tbody>
</table>

Results based on standardized variance method of identification. Bold numbers on the diagonal are estimated variances. Covariances are below the diagonal, and correlations are above. Latent means set to zero at T1 to establish metric; subsequent means therefore reflect change from T1. IN, inattention; HI, hyperactivity/impulsivity; SCT, sluggish cognitive tempo; T1, preschool; T2, post-1st grade; T3, post-4th grade; T4, post-9th grade. © 2016 Association for Child and Adolescent Mental Health.
variance) terms. Both the IN ($\chi^2(\text{df}) = 2186.1(1402); \text{CFI} = .97; \text{TLI} = .97; \text{RMSEA}[90\% \text{ CI}] = .034 [.031,.037]$) and SCT simplex models ($\chi^2(\text{df}) = 984.7 (819); \text{CFI} = .99; \text{TLI} = .98; \text{RMSEA}[90\% \text{ CI}] = .020 [.015,.025]$) provided excellent fit to the data. In comparison, the HI simplex demonstrated good fit ($\chi^2(\text{df}) = 2520.4(1402); \text{CFI} = .94; \text{TLI} = .94; \text{RMSEA}[90\% \text{ CI}] = .040 [.038,.043]$). The simplex models showed high levels of stability between the time points, with standardized transmission paths all .69 or higher. Overall, latent factors of parents’ ratings of their children’s preschool behavior were moderately correlated with the children’s behavior up to 10 years later ($r_s = .35$ for IN, .38 for HI, and .43 for SCT).

Discussion

We examined the developmental stability and temporal invariance of ADHD and SCT among a representative, community sample of 489 children across a 10-year span. The overall goals of the study were to test the developmental stability of the SCT construct and its relation with symptoms of ADHD over a 10-year period. Strong support was found for the temporal measurement invariance of these constructs – at least as assessed with parent ratings – across childhood and adolescence. While linear comparisons showed that mean levels of HI decrease with time, mean levels of SCT tend to increase slightly with age, and IN ratings are generally stable, latent simplex analyses provided support for the developmental stability of these dimensions.

Implications for the validity of SCT

The current study is the first investigation of children and adolescents to explicitly examine the measurement invariance of SCT, as well as the developmental stability and relations between ADHD and SCT dimensions for more than 2 years. The finding that like-item loadings and thresholds were found to be statistically equivalent across measurement occasions provides support for the underlying construct validity of the SCT dimension as assessed by our seven-item measure. This demonstration of measurement invariance, as well as the latent stability of parent-rated SCT throughout childhood and into adolescence, is a significant advance in evaluating the diagnostic validity of this construct (Cantwell, 1980).

Our findings of high developmental stability provide important new information regarding the ‘natural history’ of SCT, one of the key criteria that Cantwell proposed to evaluate the validity of a disorder. Further, a recent meta-analysis of extant SCT literature (Becker et al., in press) suggests that SCT is associated with multiple aspects of social and academic impairment and other psychosocial correlates, supporting the external validity of SCT as a potential diagnostic construct. However, data are sparse regarding several of the other levels of analysis described by Cantwell. Only a handful of studies has examined the external correlates of groups selected directly due to elevations of SCT (e.g., Barkley, 2013), included measures of SCT symptoms in studies of clinical intervention (e.g., Ludwig, Matte, Katz, & Rohde, 2009; Pfiffner et al., 2007), examined the etiological influences on SCT (Morruzi, Rijsdijk, & Battaglia, 2014), or explored SCT-related neurophysiology (Fassbender, Krafft, & Schweitzer, 2015). Although the current result represent an important step forward in the initial evaluation of the validity of the construct of SCT, much more work is needed before it will be possible to take a definite stance on SCT’s clinical and/or diagnostic status.

Figure 2 Latent simplex models of inattention (IN), hyperactivity/impulsivity (HI), and sluggish cognitive tempo (SCT) at preschool, post-K, post-1st, post-2nd, post-4th, and post-9th grade. For visual clarity, each dimension’s indicators are not depicted. Numbers above the endogenous latent variables indicate the percentage of variance unaccounted for in each latent variable. All path coefficients are standardized estimates, $p < .001$. © 2016 Association for Child and Adolescent Mental Health.
Longitudinal stability of ADHD and SCT

In agreement with previous studies of children with ADHD (e.g., Lahey & Willcutt, 2010), mean levels of HI symptoms in the population declined significantly across development, whereas IN remained stable. In contrast, mean levels of SCT significantly increased over this 10-year timespan. Although the current study is the first to demonstrate small but significant age-related increases in SCT in a longitudinal study of the same individuals over time, it is important to note that these small increases may reflect the development and progression of SCT psychopathology or the increasing manifestation of SCT as performance demands in the school setting increase over time.

This mean level increase in SCT is consistent with Barkley’ (2012, 2013) nationally representative cross-sectional studies that found that individuals with elevated SCT symptoms were older than individuals with elevated ADHD or co-occurring ADHD and SCT symptoms. However, other studies have not found age differences between youth with high or low levels of SCT (e.g., Becker, 2014; Lee et al., 2014). Taken together, these results suggest that additional research is needed to clarify the relation between age and SCT.

Finally, the latent correlations between the IN, HI, and SCT dimensions were relatively stable across development. The correlations, although high, were significantly less than 1.0, consistent with previous findings that SCT is separable from the ADHD dimensions (e.g., Barkley, 2013; Penny et al., 2009; Willcutt et al., 2012). This finding is also in line with a recent meta-analysis of SCT that found correlations of SCT and IN ratings of .63 for children and adolescents and .72 for adults (Becker et al., in press).

Limitations and future directions

Strengths of the current study include the use of a large community sample assessed six times over a 10-year period. The same measures of ADHD and SCT were obtained at all assessments, and the high rate of retention (approximately 92% retention through ninth grade) helped to simplify interpretation and maximize statistical power. Despite these strengths, the current study also has several important limitations that should be taken into account when interpreting the results.

Use of twins. All participants in the current study were members of twin pairs recruited through a community twin registry. Although this sampling procedure facilitated the recruitment of a sample that is generally representative of the overall population of twins in Colorado, any effects of being a twin may limit generalizability to the larger population of singletons. The stability coefficients in the current analyses are similar to the effects reported in earlier follow-up studies of nontwin samples over shorter intervals, supporting the validity of the current findings. Nonetheless, the current conclusions would be strengthened by replication in a longitudinal sample of nontwins.

Measurement of ADHD and SCT. Due to time and budgetary constraints of the overall study, SCT and DSM-IV ADHD were measured by parent ratings rather than a full-structured diagnostic interview, and teacher ratings were only obtained for a subsample of participants in two assessment waves. Secondary analyses indicated that parent–teacher inter-rater reliabilities at the end of second grade (\(n_H = .55, n_{HI} = .42, n_{SCT} = .49\); all \(p < .001\)) were comparable to those reported in a recent meta-analysis of ADHD (Willcutt et al., 2012). In addition, recent 1- and 2-year longitudinal studies of SCT using mother, father, primary teacher, and secondary teacher ratings report stability estimates that closely match those founds in the present sample, supporting the generalizability of our findings (Bernad et al., 2014, 2015). Nonetheless, future studies that include teacher ratings, as well as child self-report ratings (e.g., Becker, Luebbe, & Joyce, 2015) and a full parent interview to assess ADHD or SCT (e.g., McBurnett, 2010), would constitute a useful extension of the current results by better isolating how much of the observed developmental stability is due to the construct itself versus rater effects.

In light of a recent multi-informant scale demonstrating convergent and discriminant validity using an initial pool of 44 candidate SCT items (McBurnett et al., 2014), our particular SCT scale may also be lacking important factors of the SCT construct, such as working memory problems. To advance the search for risk factors and neuropsychological, behavioral, and neurophysiological indicators of SCT, subsequent studies of ADHD and related pathologies could easily include a brief measure of the SCT construct. More generally, the wide variability in number, type, and wording of SCT items is a pervasive limitation of the literature on SCT. The development and validation of a standard set of SCT items would significantly advance the SCT field and make findings across studies more easily comparable (see Becker et al. (in press) for a review of the most promising SCT items).

Limitations of SEM. An important limitation of any SEM approach is the existence of alternative, equally well-fitting models. The current model only provides an acceptable, parsimonious description of the data. Although the current literature strongly supports a two-factor (i.e., correlated) model of ADHD with separate IN and HI dimensions (Willcutt et al., 2012), bifactor (i.e., hierarchical) models of ADHD (e.g., Toplak et al., 2012) have also been shown to demonstrate comparable or superior fit to two-factor
models (for a review, see Willoughby & Blanton, 2015). Early evidence suggests that SCT still falls outside the umbrella of ADHD even when using bifactor models of these constructs (Garner et al., 2014; Lee, Burns, Beauchaine, & Becker, in press), but these and other alternative models warrant further exploration in future studies.

Conclusions
Results of longitudinal analyses provide the first empirical evidence that the SCT construct can be reliably measured over a 10-year period from early childhood through adolescence. These results provide important support for the construct validity of SCT and the discriminant validity of ADHD and SCT.

Supporting information
Additional Supporting Information may be found in the online version of this article:

Table S1. Stability and Intertrait Correlations of Mean Parent Ratings of IN, HI, and SCT.

Acknowledgements
This research was supported by grants from the National Institute of Child Health and Human Development (R01 HD38526 and R01 HD68728), and M. C. was supported by an institutional postdoctoral training grant (S T32 HD007289-29). The authors were also supported by NIH grants P50 HD27802 and R24 HD75460 during the preparation of this report. The authors have no competing or potential conflicts of interest.

Correspondence
Daniel Leopold, Department of Psychology and Neuroscience, UCB 345, University of Colorado, Boulder, CO 80309-0345, USA; Email: Daniel.Leopold@colorado.edu

Key points
- Sluggish cognitive tempo (SCT) represents a behaviorally defined construct characterized by inconsistent alertness and slow thinking and behavior.
- Measurement invariance and longitudinal stability are essential criteria for diagnostic validity that have been previously supported for the attention-deficit/hyperactivity disorder (ADHD) diagnosis and dimensions, but not for SCT.
- Using a longitudinal, community sample of children, the current study demonstrated temporal invariance and high stability of the ADHD and SCT dimensions between preschool and after ninth grade.
- Results provide further evidence for both the stability and separability of the ADHD and SCT dimensions, as well as the notion that SCT can be reliably measured with parent ratings at different time points during childhood and adolescence.

References

© 2016 Association for Child and Adolescent Mental Health.


Accepted for publication: 11 November 2015

© 2016 Association for Child and Adolescent Mental Health.