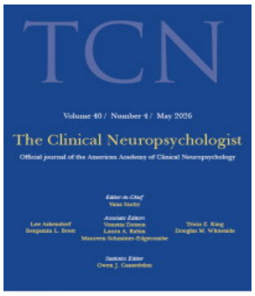

The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: www.tandfonline.com/journals/ntcn20

Intraindividual variability, how do I measure thee? Let me count the ways

Juliana Wall, Kaylee Litson, Ioannis Pavlidis & Paul T. Cirino

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2025.2601742

Intraindividual variability, how do I measure thee? Let me count the ways

Juliana Wall[a], Kaylee Litson[a], Ioannis Pavlidis[b] and Paul T. Cirino[a]

[a]Department of Psychology, University of Houston, Houston, TX, USA; [b]Department of Computer Science, University of Houston, Houston, TX, USA

ABSTRACT

Objective: The present study aimed to better understand key conceptualizations and operationalizations of intraindividual variability (IIV). We expected that differing types and metrics of IIV would relate to one another and predict outcomes (academic achievement) similarly. **Method:** The sample comprised 238 young adults. IIV was computed within and across six measures – three related to math and three more generally cognitive; in each case, score was separated from response time. We computed three types of IIV (inconsistency, dispersion, and dispersion of inconsistency), across several metrics (standard deviation, coefficient of variability, residualized standard deviation), and assessed their interrelations, and their prediction of academic achievement. **Results:** Differing metrics of variability were related to one another, but variably so. For prediction, whether or not inconsistency IIV metrics were significant was highly dependent on the measure they were derived from, with or without the primary score for a given measure also included. For dispersion of inconsistency and dispersion, variability metrics were often significant, though this was eliminated in most cases when score was also included in models. **Conclusions:** By concurrently examining multiple metrics and types of IIV within the same set of measures, this study highlights the need to (a) clarify the type of IIV utilized and why; (b) clarify the rationale for the kinds of measures used to compute IIV, particularly dispersion; and (c) include score alongside timing. Doing so will likely improve the generalizability of IIV findings, and prompt future research avenues, both psychometric- (e.g. simulations) and clinical-related (e.g. across ages and populations).

ARTICLE HISTORY
Received 2 April 2025
Accepted 4 December 2025
Published online 17 December 2025

KEYWORDS
Intraindividual variability; dispersion; inconsistency; standard deviation; coefficient of variation

Introduction

The field of neuropsychology increasingly recognizes the relevance of *variability* in patterns of performance, separate from and/or in addition to *levels* of performance. Variability can be considered relative to other individuals, or within an individual, across tests or domains, or within a single test. The overall goal of this paper is to evaluate the properties of intraindividual variability (IIV), including how it is

CONTACT Juliana Wall jwall3@uh.edu Department of Psychology, University of Houston, Houston, TX, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13854046.2025.2601742.

conceptualized and computed, and how such metrics relate to one another and predict outcomes (academic achievement in the present case).

One way that variability is considered in neuropsychology is through traditional pattern analysis, where performance level across cognitive domains is expected to produce selective weakness, depending on the known or suspected neurological disorder—in other words, expected variability in performance levels across tasks. However, variability can also be considered in its more statistical sense—the extent to which scores demonstrate a lack of consistency, either within or across tasks. Such variability is primarily ipsative, rather than gauged against norms to produce a standard score.

When viewed in this way, intraindividual variability (IIV) can be construed as an *amount* of cognitive variability. One rationale for examining differences in IIV is that with aging, speed and white matter changes produce degradations of efficiency that manifest as uneven performance (Halliday et al., 2019; Tamnes et al., 2012). Thus, variability might be expected to increase naturally with age (Bielak et al., 2014; Walhovd & Fjell, 2007), and/or in accelerated fashion with certain cognitive disorders (Haynes et al., 2017; Jackson et al., 2012). At young ages, if a neurodevelopmental disorder inhibits or otherwise alters the normal trajectory of white matter changes, this too may lead to increased variability (Kölle et al., 2022). Prefrontal networks have also been implicated in IIV (Costa et al., 2019; Kiselica et al., 2024). Such network involvement may reflect executive control failures, such an inability to allocate mental resources consistently over tasks (Penheiro et al., 2025; Vasquez et al., 2018). Even for persons with no known or suspected neurological disease, there may still be individual differences in such efficiencies (Salthouse, 2012; Schretlen et al., 2003).

The present study is grounded in the theoretical perspectives that conceptualize IIV as a marker of underlying neural and executive control processes, and more broadly, as a relatively stable individual difference rather than simply measurement error or noise (Costa et al., 2019; Nesselroade & Salthouse, 2004). Theoretical views of IIV are broad, and so might be expected to apply similarly regardless of how IIV is measured. However, very few IIV studies have directly compared *types, metrics, and definitions* of IIV (see **Glossary** in the Supplementary Appendix). Specifically, it is not always clear in the literature whether types of IIV (e.g. dispersion, inconsistency) yield similar results, or whether the metric(s) of IIV (e.g. simple standard deviation, residualized standard deviation, or coefficient of variation) should be used. It is also not clear whether the assessment of IIV should use particular kinds of tasks (e.g. from a single versus mixed cognitive domain). The present study evaluates each of the above, and in doing so can provide insight into the consistency and comparability among approaches to IIV.

Types of IIV

There are at least three types of IIV. When IIV derives variability across a variety of tasks (typically at a single time point), IIV is referred to as *dispersion* (Fellows & Schmitter-Edgecombe, 2015; Hilborn et al., 2009; Mella et al., 2016). A separate type of IIV is within-task variability, referred to as *inconsistency*. Of note, inconsistency can also refer to variability across occasions (Costa et al., 2019; Hultsch & MacDonald, 2004), but in the present study, we refer to it only within a task on a single occasion. The most common dependent variable in studies of inconsistency is response time (e.g. complex or choice reaction time). Some studies measure overall inconsistency, though others assess it across blocks/epochs of a task, or with regard to other task parameters (e.g. interstimulus interval). For example, one study found that on a continuous performance task, adolescents with ADHD exhibited greater reaction time variability than controls, with a pattern of faster reaction times during shorter interstimulus intervals (Lin et al., 2015). While studies of inconsistency are about as prevalent as those on dispersion, they examine a wider range of populations. Further, whereas most studies of dispersion are constrained to adults, studies of inconsistency are common across age ranges. In general, inconsistency follows a developmental trajectory, being highest in early childhood, declining throughout adolescence and young adulthood, and increasing again in older age (Fagot et al., 2018; Williams et al., 2005). Further, a meta-analysis found that age differences between younger and older adults were more pronounced for more demanding choice reaction time tasks than for simple reaction time tasks (Dykiert et al., 2012).

A third type of IIV, which has rarely been considered, lies at the *intersection* of dispersion and inconsistency. Rather than focusing only within a single task (as in inconsistency) or only across tasks (as in dispersion), *dispersion of inconsistency* does both—like inconsistency, it examines within-task variability, but like dispersion, it does so across tasks. To our knowledge, only one study has examined dispersion of inconsistency, finding that there were significant small to medium correlations between inconsistency and dispersion of inconsistency (Hultsch et al., 2002). This pattern generally held whether examining the entire sample or specific age groups (including younger adults). A few studies have also examined dispersion of inconsistency more indirectly than directly. For example, Epstein et al. (2011) examined inconsistency in reaction times over five different tasks, but did not directly compare or combine these across the tasks into a single index. Other studies include both dispersion and inconsistency, but as derived from different tasks. For example, Hilborn et al. (2009) assessed inconsistency *via* two reaction time tasks, and assessed dispersion across nine cognitive tasks, finding a modest correlation between these types of IIV ($0.31 \leq r \leq 0.38$). However, IIV type and measure are confounded in such a study, and so making such comparisons from the same set of tasks would be both relevant and useful.

Other methodological parameters in IIV research: metric and definition

Beyond IIV types, other key issues in IIV research include (a) how different metrics of IIV (e.g. coefficient of variation, raw standard deviation, residualized standard deviation) compare to one another, (b) how IIV is defined (which includes the content or domain of measures to be used), and (c) whether or not the actual score/performance indicator of a measure is included). The **Glossary** in the Supplementary Appendix offers further information on these terms. In terms of IIV metrics (for inconsistency), Stawski et al. (2019) noted that at least four are common: raw standard deviation (SD); coefficient of variation (CoV), which is the ratio of the standard deviation to the mean; residualized SD (residualized on mean level or other factors, resulting in a zero correlation with the respective mean level or other factors); and ex-Gaussian and/or other mathematical or mixed approaches. The authors identified the frequencies of these approaches in the literature as 24%, 30%, 29%, and 16%, respectively. They evaluated the raw SD, CoV, residualized SD, tau (an ex-Gaussian parameter that captures the tail of the response distribution), and mean squared successive differences (MSSD or variability from trial to trial). Across two separate samples of healthy older adults, within-person correlations between each of these metrics were strong for the raw SD, residualized SD, tau, and MSSD ($0.94 \leq r \leq 0.99$); however, correlations between the CoV and the other metrics were much more modest (Sample 1: $0.59 \leq r \leq 0.61$; Sample 2: $0.69 \leq r \leq 0.77$). Moreover, each of the IIV metrics except CoV significantly predicted attention switching. Although no analogous study is available for dispersion, reviews note that CoV and SD metrics are most common (e.g. Aita et al., 2024; Vance et al., 2024). Most studies do not describe a specific rationale for the metric(s) of IIV chosen. Where studies (primarily of aging) have evaluated a number of IIV metrics (e.g. Batterham et al., 2014; Lövdén et al., 2007; Stawski et al., 2019), these generally show agreement, with the exception of the CoV, which is less predictive of outcomes than other IIV metrics.

Studies also vary widely with regard to what type of task/measure content is included (IIV definitions in the parlance of this study; see **Glossary**). Some dispersion studies include only measures within a given cognitive domain; these include executive function (Buczylowska & Petermann, 2018; De Felice & Holland, 2018; Karr et al., 2014) or memory (Vandermorris et al., 2013). Many others examine dispersion across cognitive domains (Grewal et al., 2023; Halliday et al., 2018; Jones et al., 2020; Klepacz et al., 2025; Scott et al., 2023; Webber et al., 2022). Sometimes this occurs within a standard test battery (e.g. RBANS; Sullivan et al., 2018), or within the subdomains of a given test (e.g. MMSE; Kliegel & Sliwinski, 2004), and sometimes this occurs across timed and untimed measures (e.g. Bangen et al., 2019). The outcome in dispersion studies of IIV is usually some sort of functional outcome, though not typically academic achievement, which may be a by-product of the fact that most dispersion studies are with adults.

Finally, not all IIV studies include the score (the accuracy raw or standard score) in addition to the variability metric (De Frias et al., 2012), and among studies that do, results vary. For example, one study found that dispersion predicted future cognitive decline better than score (Anderson et al., 2018), while a different study found that score predicts outcomes (e.g. dementia status) better than dispersion (DesRuisseaux et al., 2025).

The present study

The purpose of this study is to provide some clarity to both the measurement and meaning of IIV by comparing several of its key conceptualizations and instantiations. In doing so, the tasks were carefully chosen. First, we only included measures that were timed, given that inconsistency is ubiquitously gauged with timed tasks. This allowed us to compute and compare all types of IIV in the most directly analogous manner possible. Also, for each task, the time taken (or number correct within a set time period) was *not* the score (the primary/typical variable of performance level).

This was done so that the "mean" value (of response times) could be differentiated from the "mean" score (accuracy or performance level). We included two kinds of tasks—those that are associated with math (domain-specific), and those that were cognitive (domain-general). The context of this study is college students; the age range is such where processing speed more generally has stabilized but is not yet declining (Kail & Salthouse, 1994; Miller & Vernon, 1997). In this sample, variability is likely to represent individual differences in IIV, rather than being amplified by either developmental factors (at either end of the age spectrum), or by cognitive processes associated with a given neuropsychological disorder.

In this study, we also include outcome measures of academic achievement (standardized measures of math computation and reading comprehension). These were chosen given that the parent study from which these data were drawn pertains to math performance in college. Also, because multiple cognitive and mathematical domains were sampled (separate from the achievement outcomes), this allowed for the association (e.g. IIV from math tasks predicting math outcomes) as well as various levels of dissociation (e.g. IIV from math tasks predicting reading outcomes; IIV from cognitive tasks predicting reading or math outcomes). Doing so allows us to address the issue of whether the kind of measure and kind of outcome influenced relations involving IIV. Aside from these practical considerations, there is also empirical precedent of a relationship between academic achievement and IIV (e.g. Borella et al., 2011; Dirk & Schmiedek, 2016; Van De Voorde et al., 2010; Wang et al., 2022). It is of note that the measures chosen all "make sense" as predictors in the academic achievement literature, even though the primary focus of this study is on the conceptualization of IIV and the extent to which different metrics of IIV similarly or dissimilarly predict outcomes.

This study provides several novel contributions. First, few prior studies directly compare dispersion and inconsistency IIV in the same sample using the same set of measures. Second, few studies have directly compared multiple metrics (e.g. mean, SD, CoV) or definitions (comparing indices derived from different combinations of tasks) of IIV. Additionally, the inclusion of dispersion of inconsistency is novel. This is the first known study to examine all of these issues concurrently.

Hypotheses

1. Consistent with the view that IIV reflects a stable marker of neural or executive control, we hypothesize that within IIV type, relations among different metrics (mean, SD, CoV) will show medium to strong correlations, regardless of the kind of measures used (definition of IIV).
2. Based on the notion that IIV reflects a construct of its own (separate from score), we expect IIV metrics (mean, SD, CoV), both within a given measure, as well as across measures, to only correlate modestly (~0.30) with score, if only because each are derived from the same sets of measures.
3. Consistent with literature showing that IIV predicts relevant outcomes, we hypothesize that IIV metrics (CoV, SD, residualized SD) will each significantly relate to academic achievement with medium effect sizes. This should hold

both within an individual measure, as well as across measures. Further, such effects should hold over and above the main performance indicator of the task.

Materials and methods

Participants

The sample consists of 238 students enrolled in math courses at either a community college or a four-year university in the southwestern US. The present study is part of a larger project, designed to evaluate predictors of college math performance. However, no other papers using this dataset focus on IIV. All students were recruited as taking their first entry level college math class (either a developmental class or credit-bearing course, or at the highest level, college algebra); students whose first college math class was advanced (e.g. Calculus) were not included. All students were aged 18–25. The sample was diverse in terms of race/ethnicity, language background, and immigrant status. Table 1 contains participant demographics.

Measures

The following measures were administered to capture IIV. The Supplementary Appendix provides a glossary of how the measures were used to produce metrics of IIV, including their acronyms and how each was operationalized.

Table 1. Participant demographics.

Variable	Statistic
Participants[A]	238
Age[B]	20.3 (1.88)
Gender	
Female	68.1
Male	29.8
Race/Ethnicity	
Asian	22.3
Black/African American	27.7
Hispanic/Latino (Any)[C]	37.4
Mixed/Other	5.0
White	7.6
Generational Status[D]	
Immigrant	26.1
First Generation	29.8
U.S. Born	44.1
Language Status[E]	
English Only	37.8
English Dom/Pri	50.8
English Not Dom/Pri	10.9
Site	
Community College	72.3
Four-Year University	27.7

Note. [A]This is a total N. [B]This is a mean and SD. All other numbers are percents. Not all percentages add to 100% given missing or combinatorial data, and/or due to rounding. [C]Includes Hipanic/Latino of any race. [D]Immigrants are individuals born outside US; First generation are individuals born in US, but not one or more parents; US are those whose immediate families (individual and parents) are both born in US. [E]English Dom/Pri means that English is dominant and/or primary language, even if an individual has a different native language and/or is fluent in another language; English Not Dom/Pri means that English is neither the dominant nor primary language spoken.

Math measures

Arithmetic concepts (Tolar, 2010, 2018). In this task, participants responded to 16 multiple-choice math questions requiring them to judge whether mathematical statements were true. No actual computations were required, and students were instructed not to compute the answer. The primary performance indicator for this task is the number of questions answered correctly. Internal consistency in this sample for all items was 0.93 for response times, and 0.83 for accuracy.

Number line tasks (fraction and decimal) (adapted from Booth et al., 2014; DeWolf et al., 2015; Hansen et al., 2015). These two tasks asked participants to estimate where a number (e.g. 0.340, 5/8) fell on a number line bounded by 0 and 2. The fraction version of this task included 25 trials, and the decimal version included 35 trials. The primary performance indicator for this task is the absolute deviation of the response from the correct location. The internal consistencies for this sample were 0.92 and 0.93 for fractions (performance and timing metrics, respectively), and 0.84 and 0.94 for decimals (performance and timing metrics, respectively).

Non-math measures

Continuous performance task (Cirino, 2022). In this experimenter-created task, stimuli were displayed in the center of a computer screen, and participants were instructed to press the spacebar as quickly as possible when a target symbol appeared. Stimuli appeared after variable interstimulus intervals (500, 1500, and 3000 ms), and the target symbol appeared on 20% of the 312 trials, divided across three blocks. Therefore, the maximum number of trials that could be included in the IIV calculation (i.e., hits) was 62. The primary performance indicator for this task is discriminability (d'), which is the z-scored difference between hits and false alarms. Internal consistency in this sample was computed across the three blocks of this task, and was 0.80 for discriminability (score), 0.90 for response time, and 0.78 for variability (standard deviation of response time). This measure was utilized in a structural attention paper with a different sample, and evidenced similar properties (Cirino, 2023).

Automated reading span (Oswald et al., 2015; Redick et al., 2012; Unsworth & Engle, 2005). In this computer task, participants were asked to recall letters while also making judgments about whether sentences made semantic sense. Participants were shown a set of between three and seven sentences, and for each of these, participants judged whether each sentence made sense or not. Following each sentence, a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) appeared on screen for 1000 ms. Participants had to recall each letter in the order that it appeared within a set. Sets were randomized in terms of order, but each participant received a total of six sets. This task is designed to assess working memory, and the score outcome is the weighted percentage of letters recalled correctly. Response times were also recorded, and IIV was computed from these response times. Internal consistency for the current sample was 0.66 for accuracy, and 0.96 for response times.

***Automated symmetry span (Foster et** al., 2015; Oswald et al., 2015; Redick et al., 2012; Unsworth & Engle, 2005). In this task, participants were asked to recall spatial information while also making perceptual judgments about whether or not patterns were symmetrical. Participants were shown a set of between two and five 8×8 grids containing a figure, and participants judged whether or not the figure was symmetrical. Following each figure, a 4×4 grid with a box highlighted in red was presented for 650 ms, and participants had to recall the position of each highlighted box within the set, in order. Like the reading span task, sets were randomized in terms of order, but each participant received a total of six sets. This task is designed to assess working memory capacity, and the primary outcome is the weighted percentage of boxes recalled correctly. Again, response times were recorded, and IIV was computed from these. Internal consistency for the current sample was 0.75 for accuracy, and 0.94 for response times.

Outcome measures

Math computations (Kaufman & Kaufman, 2014). This task, from the KTEA-3, assessed math skills such as addition, subtraction, multiplication, and division; fractions and decimals; square roots and exponents; and algebra. Items increased in difficulty as the task progressed. The participant's score was the number of correct items prior to their ceiling of four consecutive incorrect items. Mean internal consistency for this subtest is 0.95 (Kaufman & Kaufman, 2014), and in the current sample was 0.95.

Reading comprehension (Kaufman & Kaufman, 2014). In this task, also from the KTEA-3, participants were instructed to read a series of short passages and answer questions about them. This task measured participants' ability to understand and make inferences about texts they read. Passages and questions increased in difficulty as the task progressed. The participant's score was the number of correct items within the set given. Mean internal consistency for this subtest is 0.88 (Kaufman & Kaufman, 2014); internal consistency in the current sample was difficult to compute, as students received differing sets of items (depending on the setting of their basal and ceiling items, per test instructions); for the most common set (received by 87% of the sample), internal consistency was 0.75.

Procedure

The Institutional Review Board of a large southwestern university, in addition to the local community college, reviewed and approved all procedures. This study was not pre-registered prior to data collection or analysis. Trained examiners administered individual assessments at each student's respective campus, and assessments lasted approximately four hours over several sessions; all computer tasks used to derive IIV were administered during a single session on the same day. As the present study is part of a larger project, additional tasks (not relevant here) were administered. Participation was voluntary, and students received gift cards for participating. To counteract order effects, the assessment battery was administered in reverse order to about half the participants. After the conclusion of data collection, additional quality control measures were implemented (e.g. score verification, basals/ceilings where relevant, data reliability issues).

Analyses

Preliminary analyses included the evaluation of variable distributions of computed IIV metrics and relations with covariates (see Glossary in the Supplementary Appendix for how IIV metrics were operationalized). IIV standardized[1] variable distributions were generally appropriate, and all standardized values ±3 for each metric were winsorized.[2] Across all IIV scores, lower values were made to indicate better cognitive efficiency. Primary analyses were then conducted by 1) examining zero-order correlation coefficients within and across IIV type (inconsistency, dispersion of inconsistency, and dispersion), 2) fitting simple and multiple linear regression models to assess the direct effect of IIV on math computation and reading comprehension outcomes, separately, and 3) fitting multiple linear regression models to assess the partial effect of IIV while also accounting for actual scores' effects on outcomes. In all regression models, multicollinearity was checked by evaluating the variance inflation factor (VIF). These analyses were all computed first in SAS, then recomputed in R v. 4.4.3 using the stats package lm() function and organized using the sjPlot package v. 2.8.17 (Lüdecke, 2024). R^2 alongside Akaike Information Criteria (AIC), β-coefficients, and change in statistical significance were used as indices of model comparisons within task domain. Results were interpreted using common language (e.g. low/weak, medium/moderate, high/strong correlations) alongside statistical estimates to contextualize effects, in line with common reporting practices (Kelley & Preacher, 2012). Specifically, we reference Cohen's (1988) guidelines for effect sizes, where squared correlations (R^2) ~ 3%, 9%, and 25% are classified as small, medium, and large effects, respectively (corresponding to correlation coefficients of roughly .10, .30, and .50).

Because of the number of models evaluated and presented in the manuscript, a naming scheme was developed to better facilitate interpretation. These model names are used consistently in Tables 3 through 5 and refer to the following schematic. Math 1a through Math 4b are the eight specified models with math computation as the primary outcome, while models Read 1a through Read 4b are the eight models with reading comprehension as the outcome. Math 1a, 2a, 3a, and 4a are models that *only contain IIV predictors* of a math outcome, while Math 1b, 2b, 3b, and 4b *contain both IIV AND score predictors* of a math outcome. Analogously, Read 1a, 2a, 3a, and 4a contain only IIV predictors, and Read 1b, 2b, 3b, and 4b contain both IIV and score predictors. Models 1a and 1b utilize the unadjusted standard deviation (SD) as an IIV metric (whether math or reading is the outcome); models 2a and 2b utilize SD and the mean as the IIV metric; models 3a and 3b utilize the coefficient of variation as the IIV metric; and models 4a and 4b include the residualized standard deviation as the IIV metric. Models with similar *numbers* can be compared to one another (e.g. Math 1a is comparable to Math 1b) while models with similar *letters* can be contrasted against one another (e.g. Math 1a contrasted against Math 2a, Math 3a, and Math 4a). For outcomes, models that have similar letters and numbers can also be meaningfully contrasted (e.g. Math 1a contrasted against Read 1a).

Results

The correlations of key variables of interest across IIV type (inconsistency, dispersion of inconsistency, and dispersion) are provided in Table 2. Primary results for the prediction of academic achievement outcomes by metrics of response time inconsistency,

Table 2. Correlations of key inconsistency and dispersion metrics of intraindividual variability.

		Reaction time (RT)			Score (SC)	
		I-SC/DI-SC	I- M_{RT}/DI-M_{RT}	I-SD_{RT}/DI-SD_{RT}	D-M_{SC}	D-SD_{SC}
Arithmetic Concepts RT	I-M_{RT}	−0.25**				
	I-SD_{RT}	−0.17*	.81**			
	I-CoV_{RT}	.11	−0.08	.47**		
CPT RT	I-M_{RT}	.33**				
	I-SD_{RT}	.47**	.67**			
	I-CoV_{RT}	.47**	.48**	.97**		
Decimal Number Line RT	I-M_{RT}	−0.18*				
	I-SD_{RT}	−0.05	.88**			
	I-CoV_{RT}	.24**	.08	.51**		
Fraction Number Line RT	I-M_{RT}	−0.06				
	I-SD_{RT}	.10	.91**			
	I-CoV_{RT}	.38**	.29**	.63**		
Reading Span RT	I-M_{RT}	.33**				
	I-SD_{RT}	.20*	.77**			
	I-CoV_{RT}	−0.12	−0.07	.55**		
Symmetry Span RT	I-M_{RT}	.19*				
	I-SD_{RT}	.23**	.86**			
	I-CoV_{RT}	.15*	.15*	.56**		
Math RT	DI-M_{RT}	−0.21*				
	DI-SD_{RT}	−0.06	.91**			
	DI-CoV_{RT}	.33**	.13	.51**		
Cognitive RT	DI-M_{RT}	.25**				
	DI-SD_{RT}	.33**	.78**			
	DI-CoV_{RT}	.26**	.21	.73**		
All RT	DI-M_{RT}	.02				
	DI-SD_{RT}	.19*	.85**			
	DI-CoV_{RT}	.39**	.15	.62**		
Math Score	D-SD_{SC}				.36**	
	D-CoV_{SC}				.14*	.96**
Cognitive Score	D-SD_{SC}				.25**	
	D-CoV_{SC}				.02	.97**
All Score	D-SD_{SC}				.34**	
	D-CoV_{SC}				.03	.95**

Note. $^*p < .05$; $^{**}p < .001$; Values in the upper left block reflect Inconsistency relationships among response time metrics *within a given* task; values in the middle left block reflect Dispersion of Inconsistency relationships among response time metrics *across multiple* tasks; values in the lower right block reflect Dispersion relationships among performance score metrics *across multiple* tasks. On average, RT measures were comprised of individual trials (see methods for specific n-trials per task) while Score measures were comprised of 3 (Math, Cognitive) or 6 (All) tasks. CPT = continuous performance task. I-SD_{RT} = The SD of the response times across the trials *within a given* task; I-M_{RT} = The average/mean response times across the trials *within a given* task; I-$SD_{RT\text{-}RZD}$ = The SD of the response times across the trials *within a given* task, residualized from (net of) their mean (e.g. correlates zero with I-M_{RT}); I-CoV_{RT} = The Inconsistency coefficient of variation of response times for a *given* task, computed as I-SD_{RT}/I-M_{RT}; I-SC = The score (accuracy or performance level) metric of a *given* task. DI-SD_{RT} = The average/mean of the SD of the response times *across multiple* tasks; DI-M_{RT} = The average/mean response times *across multiple* tasks. DI-$SD_{RT\text{-}RZD}$ = The average/mean of the SD of the response times *across multiple* tasks, residualized from (net of) the mean (e.g. correlates zero with DI-M_{RT}). DI-CoV_{RT} = The Dispersion of Inconsistency coefficient of variation of response times *across multiple* tasks, computed as DI-SD_{RT}/DI-M_{RT}. DI-SC = The average/mean score (accuracy or performance level) *across multiple* tasks. D-SD_{SC} = The SD of the score (accuracy or performance level) *across multiple* tasks. D-M_{SC} = The average/mean of the score (accuracy or performance level) *across multiple* tasks. D-$SD_{SC\text{-}RZD}$ = The SD of the score (accuracy or performance level) *across multiple* tasks, residualized from (net of) the mean (correlates zero with D-M_{SC}). D-CoV_{SC} = The Dispersion coefficient of variation of score *across multiple* tasks, computed as D-SD_{SC}/D-M_{SC}. See Glossary for more details about terminology and their acronyms.

response time dispersion of inconsistency, and score dispersion are displayed in Tables 3, 4, and 5, respectively. VIF showed generally acceptable results, $M = 1.92$ and $SD = 1.51$.[3]

Agreement of metrics within IIV type

The first hypothesis was that the metrics of IIV (e.g. standard deviation alone, coefficient of variation) would be highly similar within a given measure. Although mean score is technically not itself a measure of variability, it is used in the computation of the coefficient of variation, and thus we expected it to also correlate with the other IIV metrics. Results variably supported the first hypothesis. As shown in Table 2 (response time inconsistency), five of the six tasks (Arithmetic Concepts, Reading Span, Symmetry Span, Number Line Decimal, and Number Line Fractions) showed a consistent pattern. For these tasks, M and SD correlated highly ($0.77 \leq r \leq 0.91$), SD and CoV moderately to highly ($0.47 \leq r \leq 0.63$), and M and CoV weakly ($-0.08 \leq r \leq 0.29$). For CPT, the pattern was different, with M and SD correlating highly ($r = 0.67$), SD and CoV near uniformity ($r = 0.97$), and mean and CoV moderately ($r = 0.48$).

The first hypothesis also stated that when each of the above metrics (mean, SD, CoV) are amalgamated across measures (i.e., dispersion of inconsistency), there would also be strong relations among these metrics. Results partially supported this prediction. As shown in Table 2, the mean and SD correlated highly ($0.78 \leq r \leq 0.91$), the SD and CoV correlated highly ($0.51 \leq r \leq 0.73$), but the mean and CoV correlated weakly ($0.13 \leq r \leq 0.21$). Results are consistent with the dominant pattern from Hypothesis 1 (within individual measures). Table 2 also demonstrates that for dispersion (of scores), CoV and SD correlated near unity ($0.95 \leq r \leq 0.97$).

The second hypothesis concerned the relation of IIV metrics to the score (see Table 2). Inconsistency and dispersion of inconsistency IIV are computed from response times. As noted, though, for the measures of this study, this is separate from the score of the measure (see Glossary). For example, the dominant indicator for performance on a working memory task is not how long they took to answer, but rather how much they remembered. Only dispersion metrics necessarily include the score in the computation of its metrics. We expected the correlations of this score variable and IIV metrics to be low (but still significantly above zero, given that all are derived from the same tasks). This was generally the case, with some exceptions. Within individual measures (first section of column 1 in Table 2; inconsistency), in only 5 of 18 cases did the score and one of the three IIV metrics correlate higher than 0.30 (a medium effect, Cohen, 1988). Three of these involved the CPT ($0.33 \leq r \leq 0.47$); the others were Fraction Number Line ($r = 0.38$ with CoV) and Reading Span ($r = 0.33$ with M). Across domains (second section of column 1 in Table 2; dispersion of inconsistency), the math score correlated moderately with math CoV ($r = 0.33$), but weaker with math mean response time ($r = -0.21$) and math response time SD ($r = -0.06$). For the cognitive domain, the score correlated similarly with all variability metrics ($0.25 \leq r \leq 0.33$). Across all measures, the score variable correlated moderately with CoV ($r = 0.39$), but weaker with mean response time ($r = 0.02$) and with response time SD ($r = 0.19$). For dispersion (bottom section of Table 2), in all cases, the score of a given

Table 3. Inconsistency intraindividual variability predicting math computation or reading comprehension.

Predictor	Math computation as outcome								Reading comprehension as outcome							
	Math 1a	Math 1b	Math 2a	Math 2b	Math 3a	Math 3b	Math 4a	Math 4b	Read 1a	Read 1b	Read 2a	Read 2b	Read 3a	Read 3b	Read 4a	Read 4b
	Arithmetic Concepts RT → Math Computation (β)								**Arithmetic Concepts RT → Reading Comprehension (β)**							
I-SD_{RT}	.10	−0.02	−0.02	.05					.11	.02	−0.11	−0.06				
I-M_{RT}			.15	−0.10							.27* [a]	.10				
I-CoV_{RT}					−0.07	−0.00							−0.09	−0.03		
I-$SD_{RT\text{-}RZD}$							−0.01	.03							−0.06	−0.03
I-SC		−0.74**		−0.75**		−0.73**		−0.74**		−0.50**		−0.49**		−0.50**		−0.50**
AIC	*1922.4*	*1746.3*	*1922.7*	*1746.7*	*1923.5*	*1746.5*	*1924.6*	*1746.1*	*1853.4*	*1790.5*	*1849.7*	*1791.6*	*1854.2*	*1790.3*	*1855.0*	*1790.3*
N	*233*	*233*	*233*	*233*	*233*	*233*	*233*	*233*	*231*	*231*	*231*	*231*	*231*	*231*	*231*	*231*
R^2 adj	***.005***	***.535***	***.008***	***.536***	***.000***	***.534***	***−0.004***	***.535***	***.007***	***.247***	***.027***	***.247***	***.003***	***.247***	***−0.000***	***.247***
	Decimal Number Line RT → Math Computation (β)								**Decimal Number Line RT → Reading Comprehension (β)**							
I-SD_{RT}	−0.15*	−0.17**	−0.53**	−0.32*					−0.06	−0.08	−0.41**	−0.22				
I-M_{RT}			.43**	.17							.39**	.16				
I-CoV_{RT}					−0.27**	−0.17**							−0.19**	−0.10		
I-$SD_{RT\text{-}RZD}$							−0.25**	−0.16**							−0.19**	−0.10
I-SC		−0.46**		−0.43**		−0.41**		−0.41**		−0.41**		−0.39**		−0.38**		−0.38**
AIC	*1932.8*	*1878.9*	*1924.5*	*1879.0*	*1920.2*	*1878.9*	*1922.7*	*1880.4*	*1865.1*	*1824.6*	*1858.9*	*1825.1*	*1857.5*	*1823.7*	*1857.1*	*1823.4*
N	*235*	*235*	*235*	*235*	*235*	*235*	*235*	*235*	*233*	*233*	*233*	*233*	*233*	*233*	*233*	*233*
R^2 adj	***.017***	***.222***	***.056***	***.225***	***.069***	***.222***	***.059***	***.217***	***−0.000***	***.163***	***.030***	***.165***	***.032***	***.166***	***.033***	***.167***
	Fraction Number Line RT → Math Computation (β)								**Fraction Number Line RT → Reading Comprehension (β)**							
I-SD_{RT}	−0.02	.04	−0.57**	−0.05					−0.08	−0.03	−0.40**	.04				
I-M_{RT}			.61**	.09							.35* [a]	−0.08				
I-CoV_{RT}					−0.25**	−0.04							−0.18**	.00		
I-$SD_{RT\text{-}RZD}$							−0.24**	−0.02							−0.17**	.02
I-SC		−0.58**		−0.57**		−0.56**		−0.57**		−0.47**		−0.49**		−0.48**		−0.48**
AIC	*1922.8*	*1830.3*	*1909.2*	*1831.9*	*1907.6*	*1830.5*	*1909.1*	*1830.8*	*1849.6*	*1793.5*	*1846.4*	*1795.2*	*1843.7*	*1793.7*	*1844.5*	*1793.6*
N	*233*	*233*	*233*	*233*	*233*	*233*	*233*	*233*	*231*	*231*	*231*	*231*	*231*	*231*	*231*	*231*
R^2 adj	***−0.004***	***.328***	***.057***	***.326***	***.059***	***.327***	***.053***	***.326***	***.002***	***.221***	***.020***	***.219***	***.028***	***.220***	***.025***	***.221***
	Continuous Performance Task RT → Math Computation (β)								**Continuous Performance Task RT → Reading Comprehension (β)**							
I-SD_{RT}	−0.03	.05	−0.08	.01					−0.07	.02	−0.10	−0.01				
I-M_{RT}			.07	.07							.04	.04				
I-CoV_{RT}					−0.06	.02							−0.09	−0.00		
I-$SD_{RT\text{-}RZD}$							−0.06	−0.01							−0.07	−0.01
I-SC		−0.19**		−0.19**		−0.17* [a]		−0.16* [a]		−0.20**		−0.20**		−0.19*		−0.18**
AIC	*1930.3*	*1925.6*	*1931.7*	*1926.8*	*1929.7*	*1926.0*	*1929.7*	*1926.1*	*1857.5*	*1852.5*	*1859.3*	*1854.2*	*1856.9*	*1852.5*	*1857.5*	*1852.5*
N	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*232*	*232*	*232*	*232*	*232*	*232*	*232*	*232*
R^2 adj	***−0.003***	***.021***	***−0.005***	***.020***	***−0.001***	***.019***	***−0.001***	***.019***	***.001***	***.027***	***−0.002***	***.024***	***.004***	***.027***	***.001***	***.027***

(*Continued*)

Table 3. Continued.

	Math computation as outcome								Reading comprehension as outcome							
Predictor	Math 1a	Math 1b	Math 2a	Math 2b	Math 3a	Math 3b	Math 4a	Math 4b	Read 1a	Read 1b	Read 2a	Read 2b	Read 3a	Read 3b	Read 4a	Read 4b
	Reading Span RT → Math Computation (β)								**Reading Span RT → Reading Comprehension (β)**							
I-SD_{RT}	−0.21**	−0.17**	−0.11	−0.16					−0.37**	−0.32**	−0.17	−0.20* [a]				
I-M_{RT}			−0.12	−0.01							−0.25**	−0.15				
I-CoV_{RT}					−0.07	−0.11							−0.07	−0.11		
I-$SD_{RT\text{-}RZD}$							−0.07	−0.10							−0.11	−0.14* [a]
I-SC		−0.23**		−0.23**		−0.27**		−0.27**		−0.25**		−0.22**		−0.33**		−0.33**
AIC	*1936.0*	*1925.4*	*1936.6*	*1927.4*	*1945.3*	*1929.0*	*1945.3*	*1929.4*	*1844.0*	*1828.9*	*1839.1*	*1828.4*	*1876.8*	*1852.1*	*1875.0*	*1850.1*
N	*236*	*236*	*236*	*236*	*236*	*236*	*236*	*236*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*
R^2 adj	***.040***	***.086***	***.041***	***.082***	***.001***	***.071***	***.001***	***.070***	***.131***	***.189***	***.152***	***.194***	***.000***	***.104***	***.008***	***.112***
	Symmetry Span RT → Math Computation (β)								**Symmetry Span RT → Reading Comprehension (β)**							
I-SD_{RT}	−0.21**	−0.15* [a]	−0.38**	−0.32*					−0.21**	−0.17**	−0.43**	−0.39**				
I-M_{RT}			.20	.20							.25	.25				
I-CoV_{RT}					−0.23**	−0.20**							−0.27**	−0.24**		
I-$SD_{RT\text{-}RZD}$							−0.19**	−0.16*							−0.22**	−0.19**
I-SC		−0.24**		−0.24**		−0.25**		−0.26**		−0.17*		−0.17*		−0.17**		−0.19**
AIC	*1898.9*	*1887.1*	*1898.6*	*1886.6*	*1896.3*	*1882.8*	*1900.5*	*1886.2*	*1828.7*	*1824.1*	*1826.9*	*1822.2*	*1821.9*	*1816.6*	*1828.3*	*1822.1*
N	*231*	*231*	*231*	*231*	*231*	*231*	*231*	*231*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*
R^2 adj	***.038***	***.090***	***.044***	***.096***	***.049***	***.107***	***.031***	***.094***	***.041***	***.064***	***.053***	***.076***	***.069***	***.094***	***.043***	***.072***

Note. $^*p < .05$; $^{**}p < .001$.

[a]= when adjusting for Bonferroni correction, this effect became non-significant; I-SD_{RT} = The SD of the response times across the trials *within a given* task; I-M_{RT} = The average/mean response times across the trials *within a given* task; I-$SD_{RT\text{-}RZD}$ = The SD of the response times across the trials *within a given* task, residualized from (net of) their mean (e.g. correlates zero with I-M_{RT}); I-CoV_{RT} = The inconsistency coefficient of variation of response times for a *given* task, computed as I-SD_{RT}/I-M_{RT}; I-SC = The score (accuracy or performance level) of a *given* task. See Glossary for more details about terminology and their acronyms. Estimates are standardized β values of regression equations. *Math 1a-Math 4b* are the eight specified models with math computation as the primary outcome. Models *Read 1a-Read 4b* are the eight models with reading comprehension as the outcome. The different model labels (1a, 1b, 2a, etc.) refer to the following sets of predictors.

1a: predictor(s) = SD 1b: Model 1a + Score.

2a: predictor(s) = SD + Mean 2b: Model 2a + Score.

3a: predictor(s) = CoV 3b: Model 3a + Score.

4a: predictor(s) = residualized-SD 4b: Model 4a + Score.

Predictor variables are presented in regular (non-italicized) font, whereas model statistics (AIC, N, R^2) are italicized to distinguish them from predictors. The R^2 values are additionally bolded to highlight them, as they represent the primary statistic of interest for model comparison and interpretation.

Table 4. Dispersion of inconsistency intraindividual variability predicting math computation or reading comprehension.

Predictor	Math 1a	Math 1b	Math 2a	Math 2b	Math 3a	Math 3b	Math 4a	Math 4b	Read 1a	Read 1b	Read 2a	Read 2b	Read 3a	Read 3b	Read 4a	Read 4b
	Math Computation as Outcome								Reading Comprehension as Outcome							
	Dispersion of Math RT Inconsistency → Math Computation (β)								**Dispersion of Math RT Inconsistency → Reading Comprehension (β)**							
DI-SD$_{RT}$	−0.02	−0.06	−0.57**	−0.05					−0.01	−0.04	−0.51**	−0.11				
DI-M$_{RT}$			.60**	−0.02							.55**	.07				
DI-CoV$_{RT}$					−0.28**	−0.05							−0.22**	−0.04		
DI-SD$_{RT\text{-}RZD}$							−0.24**	−0.02							−0.21**	−0.05
DI-SC		−0.72**		−0.72**		−0.70**		−0.71**		−0.57**		−0.56**		−0.55**		−0.55**
AIC	*1946.5*	*1778.8*	*1932.9*	*1780.8*	*1927.3*	*1779.4*	*1932.8*	*1780.5*	*1877.8*	*1788.8*	*1866.9*	*1790.5*	*1865.9*	*1788.8*	*1866.8*	*1788.7*
N	*236*	*236*	*236*	*236*	*236*	*236*	*236*	*236*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*
R^2 adj	***−0.004***	***.509***	***.056***	***.507***	***.075***	***.508***	***.053***	***.505***	***−0.004***	***.316***	***.045***	***.314***	***.045***	***.316***	***.042***	***.316***
	Dispersion of Cognitive RT Inconsistency → Math Computation (β)								**Dispersion of Cognitive RT Inconsistency → Reading Comprehension (β)**							
DI-SD$_{RT}$	−0.22**	−0.13* [a]	−0.26*	−0.17					−0.31**	−0.22**	−0.29**	−0.20* [a]				
DI-M$_{RT}$			.05	.05							−0.03	−0.03				
DI-CoV$_{RT}$					−0.18**	−0.10							−0.20**	−0.12		
DI-SD$_{RT\text{-}RZD}$							−0.16*	−0.10							−0.18**	−0.11
DI-SC		−0.30**		−0.30**		−0.32**		−0.32**		−0.28**		−0.28**		−0.32**		−0.33**
AIC	*1941.8*	*1922.6*	*1943.6*	*1924.3*	*1946.4*	*1923.9*	*1947.9*	*1924.0*	*1861.1*	*1844.6*	*1863.0*	*1846.5*	*1875.3*	*1852.6*	*1877.2*	*1853.1*
N	*237*	*237*	*237*	*237*	*237*	*237*	*237*	*237*	*235*	*235*	*235*	*235*	*235*	*235*	*235*	*235*
R^2 adj	***.045***	***.124***	***.042***	***.121***	***.027***	***.119***	***.021***	***.118***	***.093***	***.158***	***.090***	***.155***	***.037***	***.129***	***.029***	***.127***
	Dispersion of All RT Inconsistency → Math Computation (β)								**Dispersion of All RT Inconsistency → Reading Comprehension (β)**							
DI-SD$_{RT}$	−0.15*	−0.03	−0.43**	−0.04					−0.20**	−0.09	−0.44**	−0.10				
DI-M$_{RT}$			.33**	.02							.29* [a]	.01				
DI-CoV$_{RT}$					−0.26**	−0.02							−0.25**	−0.03		
DI-SD$_{RT\text{-}RZD}$							−0.22**	−0.02							−0.23**	−0.05
DI-SC		−0.64**		−0.64**		−0.64**		−0.64**		−0.54**		−0.54**		−0.54**		−0.54**
AIC	*1948.7*	*1828.9*	*1943.5*	*1830.9*	*1937.6*	*1829.1*	*1941.8*	*1829.0*	*1875.9*	*1797.2*	*1872.1*	*1799.2*	*1870.5*	*1799.6*	*1872.0*	*1799.0*
N	*237*	*237*	*237*	*237*	*237*	*237*	*237*	*237*	*235*	*235*	*235*	*235*	*235*	*235*	*235*	*235*
R^2 adj	***.017***	***.410***	***.042***	***.407***	***.062***	***.409***	***.045***	***.409***	***.034***	***.312***	***.054***	***.309***	***.056***	***.305***	***.050***	***.307***

Note. $*p < .05$; $**p < .001$.

[a]= when adjusting for Bonferroni correction, this effect became non-significant; DI-SD$_{RT}$ = The average/mean of the SD of the response times *across multiple* tasks; DI-M$_{RT}$ = The average/mean response times *across multiple* tasks. DI-SD$_{RT\text{-}RZD}$ = The average/mean of the SD of the response times *across multiple* tasks, residualized from (net of) the mean (e.g. correlates zero with DI-M$_{RT}$). DI-CoV$_{RT}$ = The Dispersion of Inconsistency coefficient of variation of response times *across multiple* tasks, computed as DI-SD$_{RT}$/DI-M$_{RT}$. DI-SC=The average/mean score (accuracy or performance level) metric *across multiple* tasks. See Glossary for more details about terminology and their acronyms. Estimates are standardized β values of regression equations. *Math 1a-Math 4b* are the eight specified models with math computation as the primary outcome. Models *Read 1a-Read 4b* are the eight models with reading comprehension as the outcome. The different model labels (1a, 1b, 2a, etc.) refer to the following sets of predictors.

1a: predictor(s) = SD 1b: Model 1a + Score.

2a: predictor(s) = SD + Mean 2b: Model 2a + Score.

3a: predictor(s) = CoV 3b: Model 3a + Score.

4a: predictor(s) = residualized-SD 4b: Model 4a + Score.

Predictor variables are presented in regular (non-italicized) font, whereas model statistics (AIC, N, R^2) are italicized to distinguish them from predictors. The R^2 values are additionally bolded to highlight them, as they represent the primary statistic of interest for model comparison and interpretation.

Table 5. Dispersion intraindividual variability predicting math computation or reading comprehension.

	Math computation as outcome						Reading comprehension as outcome					
Predictor	Math 1a	Math 1b	Math 3a	Math 3b	Math 4a	Math 4b	Read 1a	Read 1b	Read 3a	Read 3b	Read 4a	Read 4b
	Math Score Dispersion → Math Computation (β)						**Math Score Dispersion → Reading Comprehension (β)**					
D-SD_{SC}	−0.19**	.07					−0.26**	−0.06				
D-CoV_{SC}			−0.03	.07					−0.16*	−0.08		
D-$SD_{SC\text{-}RZD}$					.07	.07					−0.06	−0.06
D-M_{SC}		−0.74**		−0.72**		−0.71**		−0.54**		−0.55**		−0.56**
AIC	*1921.6*	*1765.5*	*1930.3*	*1765.5*	*1929.4*	*1765.5*	*1841.9*	*1771.7*	*1852.5*	*1770.6*	*1857.5*	*1771.7*
N	*234*	*234*	*234*	*234*	*234*	*234*	*232*	*232*	*232*	*232*	*232*	*232*
R^2 adj	***.033***	***.506***	***−0.003***	***.506***	***.001***	***.506***	***.065***	***.312***	***.021***	***.315***	***−0.000***	***.312***
	Cognitive Score Dispersion → Math Computation (β)						**Cognitive Score Dispersion → Reading Comprehension (β)**					
D-SD_{SC}	−0.08	.01					−0.06	.03				
D-CoV_{SC}			−0.00	.01					.01	.02		
D-$SD_{SC\text{-}RZD}$					.01	.01					.03	.03
D-M_{SC}		−0.35**		−0.35**		−0.35**		−0.36**		−0.35**		−0.35**
AIC	*1945.0*	*1917.9*	*1946.6*	*1917.9*	*1946.6*	*1917.9*	*1877.0*	*1848.5*	*1877.8*	*1848.6*	*1877.7*	*1848.5*
N	*236*	*236*	*236*	*236*	*236*	*236*	*234*	*234*	*234*	*234*	*234*	*234*
R^2 adj	***.002***	***.114***	***−0.004***	***.114***	***−0.004***	***.114***	***−0.000***	***.118***	***−0.004***	***.117***	***−0.004***	***.118***
	All Score Dispersion → Math Computation (β)						**All Score Dispersion → Reading Comprehension (β)**					
D-SD_{SC}	−0.17*	.06					−0.21**	−0.02				
D-CoV_{SC}			.04	.06					−0.05	−0.02		
D-$SD_{SC\text{-}RZD}$					.05	.05					−0.02	−0.02
D-M_{SC}		−0.66**		−0.65**		−0.64**		−0.55**		−0.56**		−0.56**
AIC	*1947.3*	*1827.6*	*1953.5*	*1827.2*	*1953.2*	*1827.6*	*1874.7*	*1799.7*	*1884.6*	*1799.6*	*1885.0*	*1799.7*
N	*237*	*237*	*237*	*237*	*237*	*237*	*235*	*235*	*235*	*235*	*235*	*235*
R^2 adj	***.023***	***.413***	***−0.003***	***.414***	***−0.002***	***.413***	***.039***	***.305***	***−0.002***	***.305***	***−0.004***	***.305***

Note. *$p<.05$; **$p<.001$; D-SD_{SC} = The SD of the score (accuracy or performance level) *across multiple* tasks; D-M_{SC} = The average/mean of the score (accuracy or performance level) *across multiple* tasks; D-$SD_{SC\text{-}RZD}$ = The SD of the score (accuracy or performance level) *across multiple* tasks, residualized from (net of) the mean (correlates zero with D-M_{SC}); D-CoV_{SC} = The Dispersion coefficient of variation of score *across multiple* tasks, computed as D-SD_{SC}/D-M_{SC}. See Glossary for more details about terminology and their acronyms. Estimates are standardized β values of regression equations. *Math 1a-Math 4b* are the eight specified models with math computation as the primary outcome. Models *Read 1a-Read 4b* are the eight models with reading comprehension as the outcome. The different model labels (1a, 1b, 3a, etc.) refer to the following sets of predictors.

1a: predictor(s) = SD 1b: Model 1a + Score.

2a: predictor(s) = SD + Mean*** 2b: Model 2a + Score***.

3a: predictor(s) = CoV 3b: Model 3a + Score.

4a: predictor(s) = residualized-SD 4b: Model 4a + Score.

Predictor variables are presented in regular (non-italicized) font, whereas model statistics (AIC, N, R^2) are italicized to distinguish them from predictors. The R^2 values are additionally bolded to highlight them, as they represent the primary statistic of interest for model comparison and interpretation.

***Of note, Models *Math 2a*, *Read 2a*, *Math 2b*, and *Read 2b* were not estimated for Dispersion, because D-M_{SC} is the same as "score" (whereas for Inconsistency, and Dispersion of Inconsistency, mean RT and score are different values). No effects were impacted by Bonferroni adjustments in the Dispersion models.

domain (math, cognitive, all measures) correlated more strongly with SD ($0.25 \leq r \leq 0.36$) than with CoV ($0.02 \leq r \leq 0.14$).

Prediction of outcomes by IIV type

Hypothesis 3 stated that IIV metrics across type would be significantly predictive of achievement, for individual measures (i.e., inconsistency), as well as across measures (dispersion of inconsistency and dispersion), with or without the score variable. In Tables 3, 4, and 5, Models 1a-4a *exclude* score, and Models 1b- 4b *include* the score. For these models, we examined standardized-β (and associated p-values[4]), adjusted-R^2, and AIC values to examine relationships within and across tasks.

Inconsistency IIV

For inconsistency models (Models 1a-4a; Table 3), adjusted-R^2 values were similar within tasks, with two exceptions. First, most models that included only I-SD_{RT} as a predictor had lower R^2 values than all other models. Second, the predictive relationship between Reading Span RT and reading varied across IIV metrics. For example, the I-M_{RT} metric (Model Read 2a) was significant, but the I-COV_{RT} (Model Read 3a) and the I-$SD_{RT\text{-}RZD}$ (Model Read 4a) metrics were not. Regarding β estimates of inconsistency (Table 3), models using I-SD_{RT} alone (Models 1a) generally showed weak relationships with both reading and math outcomes ($-0.21 \leq \beta \leq 0.11$), except for Reading Span I-SD_{RT} predicting reading ($\beta \leq -0.37$). β estimates of inconsistency were relatively similar across I-CoV_{RT} and I-$SD_{RT\text{-}RZD}$, yet these results differed from I-SD_{RT}, suggesting that using an unadjusted SD as a metric of IIV may yield different results than using an adjusted metric of IIV, like CoV or residualized SD. Specifically, the models that included either I-CoV or I-$SD_{RT\text{-}RZD}$ typically resulted in a larger β magnitudes for models that included math predictors (Arithmetic Concepts, Decimal Number Line, and Fraction Number Line) but similar β estimates for models that included cognitive predictors (CPT, Reading Span, Symmetry Span). In all, the hypothesis that IIV metrics relate similarly to performance scores for inconsistency IIV is only partially supported and seems to vary across task domain.

We also hypothesized that because IIV metrics are meant to be useful, then across tasks, they should predict achievement, over and above score (Models 1b through 4b, Table 3). This was not always the case, yet results were generally consistent across metrics. In Table 3, the IIV and score together robustly predicted a substantial degree of variability, especially when using specific math tasks (Arithmetic Concepts, Decimal Number Line, Fraction Number Line) to predict math outcomes ($0.22 \leq R^2 \leq 0.54$) and to a lesser degree when using specific math tasks to predict reading outcomes ($0.17 \leq R^2 \leq 0.25$). Specific cognitive tasks (CPT, Reading Span, Symmetry Span) accounted for a much lower proportion of variance when predicting math ($0.02 \leq R^2 \leq 0.11$) or reading outcomes ($0.02 \leq R^2 \leq 0.19$). These results are further supported by decreases in AIC between models *including* (models ending in b) score as a predictor versus *excluding* it (see Table 3). When considering the effect of each IIV metric in a model that also included score, score consistently and strongly predicted math and reading outcomes. Among some task domains, both score and the IIV metric predicted the outcome. For example, Decimal Number Line RT IIV metrics all predicted math even when the model included score ($-0.16 \le \beta \le -0.32$, *p*s < 0.05), and Symmetry Span RT IIV metrics all predicted reading and math even when the model included score ($-0.15 \le \beta \le -0.39$, *p*s < 0.05). Among other predictors, only some IIV metrics resulted in statistically significant β when also including score. For example, Reading Span I-SD_{RT} predicted math ($\beta = -0.17$, $p < 0.05$; Model Math 1b) above and beyond score, yet Reading Span I-CoV_{RT} and I-$SD_{RT\text{-}RZD}$ did not ($\beta = -0.07$, $\beta = -0.07$, *p*s > 0.05; Models Math 2b and 3b). Still in other measures, no IIV metrics resulted in statistically significant β when also including the score, despite IIV metrics resulting in statistically significant effects without including score (see Table 3). All four metrics (I-SD_{RT}, I-SD_{RT}+I-M_{RT}, I-CoV_{RT}, and I-$SD_{RT\text{-}RZD}$) showed similar patterns of predictions within task domains after including I-SC_{RT} as an additional predictor, with few exceptions.[5] Overall, the hypothesis that IIV metrics are useful above and beyond score is only partly supported, is relatively consistent regardless of IIV metric, and seems to vary across task domain. This highlights the importance of task selection in computing IIV, at least for the present sample and outcomes.

Dispersion of inconsistency

We purposefully selected tasks that had some bearing on our achievement outcomes (e.g. math cognition is known to be relevant for math; attention and working memory are known to be relevant for both math and reading). However, to the extent that IIV metrics represent an individual difference in their own right, then it should not matter from which task(s) they are derived. Therefore, we created dispersion metrics (mean, SD, CoV, and score) from the three math-related tasks (Arithmetic Concepts and Decimal and Fraction Number Lines) and from the three cognitive tasks (CPT, Reading Span, Symmetry Span), as well as across all variables. Table 4 shows results for the dispersion of inconsistency models. When predicting math outcomes from dispersion of inconsistency in math tasks ($0.50 \le R^2 \le 0.51$), only score was uniquely predictive ($-0.72 \le \beta \le -0.70$). The pattern was the same when predicting reading comprehension from math dispersion of inconsistency metrics ($0.31 \le R^2 \le 0.32$; $-0.57 \le \beta \le -0.54$), and when predicting combined math and reading outcomes from dispersion of inconsistency metrics in cognitive tasks ($0.12 \le R^2 \le 0.16$; $-0.33 \le \beta \le -0.28$). Overall, these findings do not support the inference that it does not matter from which task(s) IIV metrics are derived; it does matter. Specifically, when the task is directly related to the outcome, the expected relationship between IIV and outcome is anticipated to be greater and account for more variance in the outcome than when the task is not directly related to the outcome.

Further, although IIV metrics predict outcomes in most of the models that exclude score (DI-SC), adding score diminished this effect. Table 4 shows that 22 of 24 models that *exclude* score resulted in a statistically significant IIV β estimate (with lower AIC values providing additional support), yet only 3 of 24 models that *include* score did so; of these 3 models, all contained DI-SD_{RT} as a predictor (one also included DI-M_{RT}) and all were from the Dispersion of Cognitive RT Inconsistency models. Overall, results did not support the hypothesis that dispersion of inconsistency IIV metrics yields robust effects above and beyond score.

Dispersion

When predicting math outcomes from dispersion of math scores ($0.50 \leq R^2 \leq 0.51$), only score was uniquely predictive ($-0.74 \leq \beta \leq -0.71$; see Table 5). The pattern of results was the same when predicting reading from dispersion of math scores ($R^2 = 0.31$; $-0.56 \leq \beta \leq -0.54$), as well as math and reading from dispersion of cognitive scores ($0.11 \leq R^2 \leq 0.12$; $-0.36 \leq \beta \leq -0.35$). Like the results from dispersion of inconsistency models, these findings do not support the inference that it does not matter from which task(s) IIV metrics are derived; when tasks are directly related to the outcome, the relationship between IIV and outcome is greater than when the task is not directly related to the outcome. Similar to all inconsistency and dispersion of inconsistency results, including score in dispersion models impacted the relationship between each IIV metric and performance, and (lower) AIC values also support interpreting the model that includes score (see Table 5). Table 5 shows that 5 of 12 models that *exclude* score as a predictor (Models 1a, 3a, and 4a) resulted in statistically significant IIV β estimate ($-0.26 < \beta < -0.16$, $ps < 0.05$), yet 0 of 12 models that *include* score as a predictor (Models 1b, 3b, and 4b) resulted in a statistically significant IIV β estimate. Of the five models that originally resulted in statistically significant IIV β estimates, four contained D-SD_{SC} as a predictor and were from either the Math Score or All Score Dispersion domains, while the fifth contained D-CoV_{SC} as a predictor and was from the Math Score Dispersion predicting Reading Comprehension model. The IIV metric D-SD_{RT} seems to function differently with Math Scores and All Scores (combined Math and Reading Scores) relative to when derived from Cognitive Scores. For this sample and for these math and reading outcomes, dispersion metrics do not seem clearly useful above and beyond score.

The above results do not consider multiple comparisons. However, across all IIV types, post-hoc analyses that adjusted for multiple comparisons *via* Bonferroni correction minimally impacted results (see Tables 3, 4, and 5).

Discussion

The goal of the present study was to better understand the similarities and differences among types, metrics, and definitions (see **Glossary**) of intraindividual variability (IIV). Given the number of procedures, a brief recap of the results is provided. Hypotheses 1 and 2 posed that within type of IIV, differing metrics would relate to one another, and modestly with score. For inconsistency IIV (see Table 2), for five of the six tasks (excepting the continuous performance task), mean response time and SD (of response times) metrics were strongly related; the coefficient of variation (CoV) and SD were moderately related, and the CoV and mean response time were weakly related. Variability metrics in general correlated weakly with the indicator of score. The pattern was similar for dispersion of inconsistency IIV. For dispersion, however, CoV and SD correlated near unity, whereas score correlated moderately with SD, and weakly with CoV. Hypothesis 3 concerned the prediction of academic outcomes. For inconsistency, whether or not its metrics were significant was highly dependent on the measure they were derived from, with or without score also included. For dispersion of inconsistency and dispersion, variability metrics were often significant, though this was eliminated in most cases when score was also included in models.

Thus, the present study found mixed support for the key hypotheses, and adds nuance to the understanding of IIV. Some metrics of IIV correlated quite strongly with one another, whereas others correlated weakly. Further, mean response time and score correlated with variability metrics to differing degrees, across types of IIV, suggesting that they are not interchangeable. When differing metrics are used for prediction, there is broad similarity in whether or not they are significantly predictive of achievement. Again, however, the results were not identical, suggesting a lack of interchangeability.

Concerning IIV type and definition (content), some measures of variability were consistently predictive of academic achievement. For inconsistency, metrics of IIV from most individual measures were predictive of outcomes, even in the context of mean (response time) also in the model, though with small effect sizes. Further, for both working memory measures predicting either math or reading, effects were seen even in the context of the score indicator, suggesting that inconsistency IIV from such tasks is relevant. However, for measures that are more directly related to math (e.g. Number Line Estimation and Arithmetic Concepts), score is more relevant. This may extend to other domains as well, and should be evaluated in future studies. For dispersion of inconsistency (amalgamated by domain), results were similar, in that variability metrics on their own were predictive of outcomes generally, but only some cognitive measures of variability remained predictive in the presence of the performance indicator. The pattern was yet different for dispersion, where there were only a few instances of variability metrics predicting outcomes (typically SD when considered alone), but all of which no longer remained significant after including score. This pattern of results highlights that (a) inconsistency, dispersion of inconsistency, and dispersion are potentially different constructs, with differing correlates; (b) variability appears to be most relevant when it is computed from response times; (c) dispersion metrics appear less likely than other IIV types to be predictive of outcomes; and (d) definitional issues (content of the measures from which IIV is derived) matters.

The results of this study have several implications for IIV research, with the acknowledgment that patterns of intraindividual variability may differ across populations not represented in this study. First, as others have recommended (e.g. DesRuisseaux et al., 2025; Vance et al., 2022), we advocate for clarity around what type of IIV is used and how it is operationalized. Unlike the above studies, however, the present recommendation is supported by direct comparisons between multiple types, metrics, and definitions of IIV in the same sample. To the extent that IIV is expected to reflect cognitive inefficiencies, it would seem that inconsistency (or dispersion of inconsistency) would be more likely to capture such momentary breakdowns, relative to dispersion, which is required to play out across whole test batteries. If only one type of IIV is used, then there should be strong rationale (other than convenience/availability) for why alternatives are not also being considered. A strong version of this line of thought might suggest that if inconsistency and dispersion are meant to operationalize the same conceptual factor (IIV), then using both together is one way of providing evidence for that conjecture. If, on the other hand, inconsistency and dispersion are meant to be different constructs, then the conditions under which one or the other would be expected to show effects should be clearly formalized. In either case then, it seems beneficial to include both types of IIV.

Second, there were predictive differences when different metrics of IIV were utilized, and no single metric consistently predicted outcomes better than others. Thus, using multiple metrics of variability is better than using a single metric. Even in cases where both the mean response time and SD are considered together (directly, *via* CoV, or *via* residualization), results were not identical. We suggest that using the mean response time and SD as separate variables (though with both included in prediction models) is a more direct way to consider both (e.g. relative to using CoV), particularly for dispersion IIV, with the added advantage that the relative contributions of these variables can be identified. This is further supported by the fact that these variables are not highly related to one another (see Table 2). At the least, when only the CoV is utilized, it would be helpful to test whether any such effects remain when the mean response time (and/or score) and SD are included as separate predictors in the same model, to help distinguish effects of variability from effects of overall performance. We could not rule out that post-combination adjustments (e.g. needing to winsorize some values of the CoV metric to meet distributional requirements) account for the differences, but if so, this would further support the use of the mean response time and SD directly rather than CoV. Future work, such as simulation studies, would be needed to establish the robustness of this finding, or suggest conditions under which one or another metric would be preferable (or where it may not matter).

A third point concerns the necessity of including score indicators alongside variability. In many studies of inconsistency IIV, the speed/timing of trials *is* the performance indicator itself (Ghisletta et al., 2018; Moy et al., 2011), versus the use of a CPT or other measures where timing and accuracy are clearly separable (Collins et al., 2018). In cases of dispersion per se, the score indicator is captured more or less by definition (though see caveats above regarding CoV). However, in situations where that is not the case (e.g. in tasks where trial-over-trial timing is measured and accuracy is also available), it seems necessary to do so—that is, to also include score. In the present study, as noted, metrics of variability were often only related to outcomes when the key performance indicator was *not* also included. If the variability metrics are predictive even in the context of this “strong covariate” (as it was for some of the inconsistency and dispersion of inconsistency metrics in this study), then it is more likely to be generalizable/replicable. This is important to mention, even though it may not apply to all the ways that IIV is utilized in the literature.

Fourth, because definitions (task content) do appear to matter, we advocate for a strong rationale for which tasks are included in the computation of IIV. On the one hand, it makes sense to derive IIV from measures of constructs that have a strong theoretical relationship to the outcome under consideration. On the other hand, doing so should increase the relevance of the *score* indicator, which could make it *more* difficult for the variability metric to contribute additional variance. In corollary, it may also be the case that stimulus format (e.g. auditory versus visual, verbal versus spatial) and/or other methodological features (e.g. the use of performance measures versus rating scales, manual versus verbal output, timed versus untimed), may alter the extent to which variability impacts outcomes when combined, in as yet unknown ways. More broadly, if tasks from different cognitive domains (e.g. language, memory, attention, executive function) are to be amalgamated, then there should also be a strong rationale for why a given combination is most salient (again other than availability or convenience).

The above suggestions may add clarity to research that uses IIV, which in turn may increase the chance that where effects are found, they are more likely to be robust and generalizable. Ironically, there is quite some variability in the way that IIV is conceptualized and operationalized, and it is not clear to what extent this variability contributes to a lack of clear findings in the literature. This is particularly important because tasks used for inconsistency are often experimental tasks rather than normed; even though tasks themselves used in the calculation of dispersion are often normed, the variability metrics themselves are not. A recent study developed norms for dispersion across a neuropsychological battery (Kiselica et al., 2024), and it is likely that as measures of neuropsychological performance increasingly become computer or online-administered, more indices of variability may become available, as they can be automatically computed (notwithstanding the timing issues that may arise, or the difficulty in amalgamating any such metrics across measures and platforms). In corollary, it also seems important to understand the psychometric properties of measures of IIV more generally. Using different metrics or different combinations of variables, as in the present study, could be construed as an index of "alternate-form" reliability. Further, longitudinal studies could provide indices of "test-retest" reliability for these metrics. This is relevant, even if individual measures have strong internal consistency reliability. For example, inconsistency can still occur within a task even where there is trial-over-trial correlation; for dispersion, even when highly reliable individual tasks are used, an internal consistency reliability metric can be computed for all the items across all tasks, but this is not the same as the reliability of the IIV metric.

Limitations

The strength of this study lies in the ability to disaggregate a number of potentially confusable issues that may arise from the IIV literature. But as with all studies, the present study has some limitations. We are aware of the fact that our study did not evaluate the metrics of variability against typical outcomes used in studies of IIV (e.g. daily living skills, medication adherence, health outcomes), and did not systematically evaluate sociodemographic or medical/disease covariates/factors. Certainly, inclusion of outside factors and other outcomes could have influenced the relation of IIV to outcomes, but the primary goal of this paper was to use these as a lens through which to examine the properties of IIV itself. Also, CoV of dispersion was not predictive in the present study, nor was the combination of score and variability (separably but in the same model). It is possible that there may have been differences in these approaches if variability alone (or CoV) was predictive of the outcomes of this study.

An additional limitation is the exclusive use of timed, computerized tasks, which could limit generalizability, as most studies of dispersion include untimed measures. However, many studies do not include *solely* untimed measures. For instance, the Trail Making Test is commonly included as a measure in studies of dispersion (e.g. Fellows & Schmitter-Edgecombe, 2015; Halliday et al., 2018), with time to completion being the primary score indicator. In this study, we chose only measures that we timed, because including untimed measures would have prevented the calculation of inconsistency and dispersion of inconsistency, which in turn would have prevented a comparison among types of IIV.

Another potential limitation might be that we used a relatively unselected college sample. Nonetheless, we feel confident in the generalization of our results and the applicability of our conclusions. A critique of the present work might suggest that IIV is only relevant in a population with known or suspected neurological or developmental conditions (or aging), given that IIV is thought to assess cognitive inefficiency. This would imply that only variability due to a disease process is relevant. But this in turn implies that variability can be segregated into that which is systematic and due to disease, that which is systematic but due to individual differences, and that which is due to random error. The latter could be approximated through test-retest or alternate forms reliability of the variability itself (as noted above). But to parse systematic variance, it would require knowing what a “normal” amount of variability is (for the particular measures being utilized and amalgamated) in the healthy population, and “subtract” that from the variability in a patient population. We are only aware of one study that included patients and controls that showed an interaction (i.e., dispersion IIV only predicted cognitive decline in a group with Parkinson’s, not in the control group; Jones et al., 2020); in this case, the patient group performed lower than controls in absolute performance, but simultaneously had more variability. This is unusual in the general population. For example, in the IQ literature, *higher* performance is associated with greater scatter across subtests (Gontkovsky et al., 2022; Hurks et al., 2013). Additionally, across neuropsychological batteries and normative samples, it has been shown that most adults have at least one score that is more than one standard deviation below their mean score, and that the more tests are administered, the more low scores an individual is likely to have (Binder et al., 2009).

Conclusions

Differing metrics of variability (standard deviation, coefficient of variation, residualized standard deviation) were related to one another, and variably to mean response time, and to score. However, for prediction, score was clearly the dominant predictor of academic achievement (where it was included in models); nonetheless, there were cases when variability metrics (typically of inconsistency) were relevant to achievement, even in that stringent context. By concurrently examining multiple metrics of dispersion, inconsistency, and dispersion of inconsistency IIV within the same set of measures, this study highlights the need (a) to clarify the type of IIV utilized and why; (b) to clarify rationale for the kinds of measures included in the computation of IIV, particularly dispersion; and (c) to include score alongside response time variability metrics. Doing so will likely improve the extent to which IIV findings are generalizable, as well as prompt future productive research avenues that are both psychometric- (e.g. simulations, reliability) and clinical-related (e.g. comparative approaches across ages and populations).

Notes

1. All Mean and SD scores are mean standardized within sample N ~ (0, 1). CoV was first computed from raw Mean and SD values, then standardized within sample N ~ (0, 1).
2. 4/5520 values were winsorized for inconsistency metrics. 32/2760 values were winsorized for dispersion of inconsistency metrics. 13/2070 values were winsorized for dispersion metrics (total values winsorized, 0.67%).

3. Across all 108 models computed, only 8 models resulted in VIF values above 5. These included Table 3's *Fraction Number Line* Math 2a, Math 2b, Read 2a, and Read 2b, as well as Table 4's *Dispersion of Math RT Inconsistency* Math 2a, Math 2b, Read 2a, and Read 2b. Otherwise, VIF ranged from 1.00 to 4.87. VIF was only present and high among those models that included both Mean and SD, as expected.
4. Original *p*-values using no correction are shown and primarily interpreted throughout the manuscript. Tables note which effects changed from statistically significant to non-significant after using the Bonferroni correction adjustment. Generally, the correction adjustment did not impact the interpreted results.
5. Results differences within domains, where IIV metrics produced differential statistically significant vs non-significant results included the following. For Reading Span RT → Math Computation: I-SD_{RT} of Reading Span retained a significant relationship with Math Computation after including I-SC_{RT} in the model, yet among all other metrics, this effect became non-significant after including I-SD_{RT} in the model. For Reading RT → Reading Comprehension: When including I-SC_{RT} in the model that included I-SD_{RT}+I-M_{RT} as predictors, I-SD_{RT} became statistically significant after previously being non-significant. This same pattern emerged for I-$SD_{RT\text{-}RZD}$, but not for I-CoV_{RT} nor I-SD_{RT}.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported in part by National Science Foundation under Grant No. 1760760, awarded to the University of Houston. The views expressed herein are those of the author alone.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aita, S. L., Del Bene, V. A., Knapp, D. L., Demming, C. E., Ikonomou, V. C., Owen, T., Campbell, I. A., Wagaman, B. N., Borgogna, N. C., Caron, J. E., Roth, R. M., & Hill, B. D. (2024). Cognitive intra-individual variability in cognitively healthy APOE ε4 carriers, mild cognitive impairment, and Alzheimer's disease: A meta-analysis. *Neuropsychology Review*. Advance online publication. https://doi.org/10.1007/s11065-024-09654-2

Anderson, A. E., Jones, J. D., Thaler, N. S., Kuhn, T. P., Singer, E. J., & Hinkin, C. H. (2018). Intraindividual variability in neuropsychological performance predicts cognitive decline and death in HIV. *Neuropsychology*, *32*(8), 966–972. https://doi.org/10.1037/neu0000482

Bangen, K. J., Weigand, A. J., Thomas, K. R., Delano-Wood, L., Clark, L. R., Eppig, J., Werhane, M. L., Edmonds, E. C., & Bondi, M. W. (2019). Cognitive dispersion is a sensitive marker for early neurodegenerative changes and functional decline in nondemented older adults. *Neuropsychology*, *33*(5), 599–608. https://doi.org/10.1037/neu0000532

Batterham, P. J., Bunce, D., Mackinnon, A. J., & Christensen, H. (2014). Intra-individual reaction time variability and all-cause mortality over 17 years: A community-based cohort study. *Age and Ageing*, *43*(1), 84–90. https://doi.org/10.1093/ageing/aft116

Bielak, A. A. M., Cherbuin, N., Bunce, D., & Anstey, K. J. (2014). Intraindividual variability is a fundamental phenomenon of aging: Evidence from an 8-year longitudinal study across young, middle, and older adulthood. *Developmental Psychology*, *50*(1), 143–151. https://doi.org/10.1037/a0032650

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *24*(1), 31–46. https://doi.org/10.1093/arclin/acn001

Booth, J. L., Newton, K. J., & Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, *118*, 110–118. https://doi.org/10.1016/j.jecp.2013.09.001

Borella, E., Chicherio, C., Re, A. M., Sensini, V., & Cornoldi, C. (2011). Increased intraindividual variability is a marker of ADHD but also of dyslexia: A study on handwriting. *Brain and Cognition*, *77*(1), 33–39. https://doi.org/10.1016/j.bandc.2011.06.005

Buczylowska, D., & Petermann, F. (2018). Intraindividual variability in executive function performance in healthy adults: Cross-sectional analysis of the NAB Executive Functions Module. *Frontiers in Psychology*, *9*, 329. https://doi.org/10.3389/fpsyg.2018.00329

Cirino, P.T. (2022). Continuous performance test [Unpublished measure]. Available from Authors.

Cirino, P. T., Farrell, A. E., Barnes, M. A., & Roberts, G. J. (2023). An evaluation of the structure of attention in adolescence. *Developmental Neuropsychology*, *48*(4), 162–185. https://doi.org/10.1080/87565641.2023.2213789

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Collins, B., Widmann, G., & Tasca, G. A. (2018). Effectiveness of intraindividual variability in detecting subtle cognitive performance deficits in breast cancer patients. *Journal of the International Neuropsychological Society: JINS*, *24*(7), 724–734. https://doi.org/10.1017/S1355617718000309

Costa, A. S., Dogan, I., Schulz, J. B., & Reetz, K. (2019). Going beyond the mean: Intraindividual variability of cognitive performance in prodromal and early neurodegenerative disorders. *The Clinical Neuropsychologist*, *33*(2), 369–389. https://doi.org/10.1080/13854046.2018.1533587

De Felice, S., & Holland, C. A. (2018). Intra-individual variability across fluid cognition can reveal qualitatively different cognitive styles of the aging brain. *Frontiers in Psychology*, *9*, 1973. https://doi.org/10.3389/fpsyg.2018.01973

De Frias, C. M., Dixon, R. A., & Camicioli, R. (2012). Neurocognitive speed and inconsistency in Parkinson's disease with and without incipient dementia: An 18-month prospective cohort study. *Journal of the International Neuropsychological Society: JINS*, *18*(4), 764–772. https://doi.org/10.1017/S1355617712000422

DesRuisseaux, L. A., Guevara, J. E., & Duff, K. (2025). Examining the stability and predictive utility of across- and within-domain intra-individual variability in mild cognitive impairment. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *40*(1), 1–12. https://doi.org/10.1093/arclin/acae054

DeWolf, M., Bassok, M., & Holyoak, K. J. (2015). From rational numbers to algebra: Separable contributions of decimal magnitude and relational understanding of fractions. *Journal of Experimental Child Psychology*, *133*, 72–84. https://doi.org/10.1016/j.jecp.2015.01.013

Dirk, J., & Schmiedek, F. (2016). Fluctuations in elementary school children's working memory performance in the school context. *Journal of Educational Psychology*, *108*(5), 722–739. https://doi.org/10.1037/edu0000076

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Age differences in intra-individual variability in simple and choice reaction time: Systematic review and meta-analysis. *PloS One*, *7*(10), e45759. https://doi.org/10.1371/journal.pone.0045759

Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., Brinkman, W. B., Froehlich, T., Simon, J. O., & Altaye, M. (2011). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. *Neuropsychology*, *25*(4), 427–441. https://doi.org/10.1037/a0022155

Fagot, D., Mella, N., Borella, E., Ghisletta, P., Lecerf, T., & De Ribaupierre, A. (2018). Intra-individual variability from a lifespan perspective: A comparison of latency and accuracy measures. *Journal of Intelligence*, *6*(1), 16. https://doi.org/10.3390/jintelligence6010016

Fellows, R. P., & Schmitter-Edgecombe, M. (2015). Between-domain cognitive dispersion and functional abilities in older adults. *Journal of Clinical and Experimental Neuropsychology*, *37*(10), 1013–1023. https://doi.org/10.1080/13803395.2015.1050360

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7

Ghisletta, P., Renaud, O., Fagot, D., Lecerf, T., & De Ribaupierre, A. (2018). Age and sex differences in intra-individual variability in a simple reaction time task. *International Journal of Behavioral Development*, *42*(2), 294–299. https://doi.org/10.1177/0165025417739179

Gontkovsky, S. T., Kreiner, D. S., Ryan, J. J., Golden, C. J., & Teichner, G. (2022). Relative subtest scatter on the WAIS-IV in a clinical sample referred for outpatient neuropsychological evaluation. *The Clinical Neuropsychologist*, *36*(8), 2205–2220. https://doi.org/10.1080/13854046.2021.1975827

Grewal, K. S., O'Connell, M. E., Kirk, A., MacDonald, S. W. S., & Morgan, D. (2023). Intraindividual variability measured with dispersion across diagnostic groups in a memory clinic sample. *Applied Neuropsychology. Adult*, *30*(6), 639–648. https://doi.org/10.1080/23279095.2021.1970552

Halliday, D., Stawski, R., Cerino, E., DeCarlo, C., Grewal, K., & MacDonald, S. (2018). Intraindividual variability across neuropsychological tests: Dispersion and disengaged lifestyle increase risk for Alzheimer's disease. *Journal of Intelligence*, *6*(1), 12. https://doi.org/10.3390/jintelligence6010012

Halliday, D. W. R., Gawryluk, J. R., Garcia-Barrera, M. A., & MacDonald, S. W. S. (2019). White matter integrity is associated with intraindividual variability in neuropsychological test performance in healthy older adults. *Frontiers in Human Neuroscience*, *13*, 352. https://doi.org/10.3389/fnhum.2019.00352

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development*, *35*, 34–49. https://doi.org/10.1016/j.cogdev.2015.02.001

Haynes, B. I., Bauermeister, S., & Bunce, D. (2017). A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia, and mortality. *Journal of the International Neuropsychological Society: JINS*, *23*(5), 431–445. https://doi.org/10.1017/S1355617717000236

Hilborn, J. V., Strauss, E., Hultsch, D. F., & Hunter, M. A. (2009). Intraindividual variability across cognitive domains: Investigation of dispersion levels and performance profiles in older adults. *Journal of Clinical and Experimental Neuropsychology*, *31*(4), 412–424. https://doi.org/10.1080/13803390802232659

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*(2), P101–P115. https://doi.org/10.1093/geronb/57.2.P101

Hultsch, D. F., & MacDonald, S. W. S. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. Dixon, L. Backman, & L.-G. Nilsson (Eds.), *New Frontiers in Cognitive Aging* (pp. 65–88). Oxford University Press. https://doi.org/10.1093/acprof:Oso/9780198525691.003.0004

Hurks, P. P. M., Hendriksen, J. G. M., Dek, J. E., & Kooij, A. P. (2013). Normal variability of children's scaled scores on subtests of the Dutch Wechsler preschool and primary scale of intelligence – Third edition. *The Clinical Neuropsychologist*, *27*(6), 988–1003. https://doi.org/10.1080/13854046.2013.797502

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease. *Neuropsychologia*, *50*(3), 357–366. https://doi.org/10.1016/j.neuropsychologia.2011.11.024

Jones, J. D., Burroughs, M., Apodaca, M., & Bunch, J. (2020). Greater intraindividual variability in neuropsychological performance predicts cognitive impairment in de novo Parkinson's disease. *Neuropsychology*, *34*(1), 24–30. https://doi.org/10.1037/neu0000577

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199–225. psyh. https://doi.org/10.1016/0001-6918(94)90003-5

Karr, J. E., Garcia-Barrera, M. A., & Areshenkoff, C. N. (2014). Executive functions and intraindividual variability following concussion. *Journal of Clinical and Experimental Neuropsychology*, *36*(1), 15–31. https://doi.org/10.1080/13803395.2013.863833

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman test of educational achievement* (3rd ed.). NCS Pearson.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152. https://doi.org/10.1037/a0028086

Kiselica, A. M., Kaser, A. N., Weitzner, D. S., Mikula, C. M., Boone, A., Woods, S. P., Wolf, T. J., & Webber, T. A. (2024). Development and validity of norms for cognitive dispersion on the Uniform Data Set 3.0 Neuropsychological Battery. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *39*(6), 732–746. https://doi.org/10.1093/arclin/acae005

Kliegel, M., & Sliwinski, M. (2004). MMSE cross-domain variability predicts cognitive decline in centenarians. *Gerontology*, *50*(1), 39–43. https://doi.org/10.1159/000074388

Klepacz, L. M., Cerino, E. S., & Hamm, J. M. (2025). Characterizing cognitive dispersion and its correlates across the adult lifespan in MIDUS. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *80*(6), gbae201. https://doi.org/10.1093/geronb/gbae201

Kölle, M., Mackert, S., Heckel, K., Philipsen, A., Ulrich, M., & Grön, G. (2022). Lower fractional anisotropy of the corticothalamic tract and increased response time variability in adult patients with ADHD. *Journal of Psychiatry & Neuroscience: JPN*, *47*(2), E99–E108. https://doi.org/10.1503/jpn.210135

Lin, H. -Y., Hwang-Gu, S. -L., & Gau, S. S. -F. (2015). Intra-individual reaction time variability based on ex-Gaussian distribution as a potential endophenotype for attention-deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica*, *132*(1), 39–50. https://doi.org/10.1111/acps.12393

Lövdén, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*(12), 2827–2838. https://doi.org/10.1016/j.neuropsychologia.2007.05.005

Lüdecke, D. (2024). sjPlot: Data visualization for statistics in social science. R package version 2.8.17. https://CRAN.R-project.org/package=sjPlot

Mella, N., Fagot, D., & De Ribaupierre, A. (2016). Dispersion in cognitive functioning: Age differences over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, *38*(1), 111–126. https://doi.org/10.1080/13803395.2015.1089979

Miller, L. T., & Vernon, P. A. (1997). Developmental changes in speed of information processing in young children. *Developmental Psychology*, *33*(3), 549–554. psyh. https://doi.org/10.1037/0012-1649.33.3.549

Moy, G., Millet, P., Haller, S., Baudois, S., De Bilbao, F., Weber, K., Lövblad, K., Lazeyras, F., Giannakopoulos, P., & Delaloye, C. (2011). Magnetic resonance imaging determinants of intraindividual variability in the elderly: Combined analysis of grey and white matter. *Neuroscience*, *186*, 88–93. https://doi.org/10.1016/j.neuroscience.2011.04.028

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *59*(2), P49–P55. https://doi.org/10.1093/geronb/59.2.P49

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Penheiro, R., Kiselica, A. M., Webber, T. A., Kamar, J. L. T., & Woods, S. P. (2025). Banking on variability: A conceptual replication of the association between cognitive dispersion and financial management. *The Clinical Neuropsychologist*. Advance online publication. https://doi.org/10.1080/13854046.2025.2522249

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks.

European Journal of Psychological Assessment, *28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Salthouse, T. A. (2012). Psychometric properties of within-person across-session variability in accuracy of cognitive performance. *Assessment*, *19*(4), 494–501. https://doi.org/10.1177/1073191112438744

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society: JINS*, *9*(6), 864–870. https://doi.org/10.1017/S1355617703960061

Scott, B. M., Austin, T., Royall, D. R., & Hilsabeck, R. C. (2023). Cognitive intraindividual variability as a potential biomarker for early detection of cognitive and functional decline. *Neuropsychology*, *37*(1), 52–63. https://doi.org/10.1037/neu0000867

Stawski, R. S., MacDonald, S. W. S., Brewster, P. W. H., Munoz, E., Cerino, E. S., & Halliday, D. W. R. (2019). A comprehensive comparison of quantifications of intraindividual variability in response times: A measurement burst approach. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *74*(3), 397–408. https://doi.org/10.1093/geronb/gbx115

Sullivan, K. L., Woods, S. P., Bucks, R. S., Loft, S., & Weinborn, M. (2018). Intraindividual variability in neurocognitive performance is associated with time-based prospective memory in older adults. *Journal of Clinical and Experimental Neuropsychology*, *40*(7), 733–743. https://doi.org/10.1080/13803395.2018.1432571

Tamnes, C. K., Fjell, A. M., Westlye, L. T., Østby, Y., & Walhovd, K. B. (2012). Becoming consistent: Developmental reductions in intraindividual variability in reaction time are related to white matter integrity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(3), 972–982. https://doi.org/10.1523/JNEUROSCI.4779-11.2012

Tolar, T.D. (2010). Arithmetic concepts test [Unpublished measure]. Available from Author.

Tolar, T.D. (2018). Test of arithmetic concepts [Unpublished measure]. Available from Author.

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, *33*(2), 213–220. https://doi.org/10.3758/BF03195310

Walhovd, K. B., & Fjell, A. M. (2007). White matter volume predicts reaction time instability. *Neuropsychologia*, *45*(10), 2277–2284. https://doi.org/10.1016/j.neuropsychologia.2007.02.022

Webber, T. A., Kiselica, A. M., Mikula, C., & Woods, S. P. (2022). Dispersion-based cognitive intra-individual variability in dementia with Lewy bodies. *Neuropsychology*, *36*(8), 719–729. https://doi.org/10.1037/neu0000856

Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology*, *19*(1), 88–96. https://doi.org/10.1037/0894-4105.19.1.88

Van De Voorde, S., Roeyers, H., Verté, S., & Wiersema, J. R. (2010). Working memory, response inhibition, and within-subject variability in children with attention-deficit/hyperactivity disorder or reading disorder. *Journal of Clinical and Experimental Neuropsychology*, *32*(4), 366–379. https://doi.org/10.1080/13803390903066865

Vance, D. E., Del Bene, V. A., Frank, J. S., Billings, R., Triebel, K., Buchholz, A., Rubin, L. H., Woods, S. P., Li, W., & Fazeli, P. L. (2022). Cognitive intra-individual variability in HIV: An integrative review. *Neuropsychology Review*, *32*(4), 855–876. https://doi.org/10.1007/s11065-021-09528-x

Vance, D. E., Xu, Y., Dastgheyb, R., Maki, P. M., Zhang, J., Springer, G., Anastos, K., Gustafson, D. R., Weber, K. M., Dykxhoorn, D. M., Milam, J., Diaz, M. M., Kassaye, S. G., Waldrop, D., Lee, J., Kempf, M.-C., Konkle-Parker, D., Goodkin, K., Leviner, A. J., … Rubin, L. H. (2024). Does cognitive intra-individual variability predict change in everyday functioning performance in women with and without HIV in the Women's Interagency HIV Study? *Applied Neuropsychology. Adult*. Advance online publication. https://doi.org/10.1080/23279095.2024.2444573

Vandermorris, S., Murphy, K. J., & Troyer, A. K. (2013). Age-related elevations in intraindividual variability on associative memory tasks. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *20*(6), 722–734. https://doi.org/10.1080/13825585.2013.772557

Vasquez, B. P., Tomaszczyk, J. C., Sharma, B., Colella, B., & Green, R. E. A. (2018). Longitudinal recovery of executive control functions after moderate-severe traumatic brain injury: Examining trajectories of variability and ex-Gaussian parameters. *Neurorehabilitation and Neural Repair*, *32*(3), 191–199. https://doi.org/10.1177/1545968318760727

Wang, Z., Sun, Y., Wang, H., Liu, Q., & Zhang, R. (2022). Intraindividual reaction time variability and academic achievement of primary school students: A longitudinal cross-lagged panel model. *Journal of Applied Developmental Psychology*, *82*, 101454. https://doi.org/10.1016/j.appdev.2022.101454