Development and Validation of Norm-Referenced Measures of Reaction Time Inconsistency

by

Paul W. H. Brewster

B. A., York University (2009)
MSc, University of Victoria (2011)

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

Supervisory Committee

Development and Validation of Norm-Referenced Measures of Reaction Time Inconsistency

by

Paul W. H. Brewster

BA, York University (2009)
MSc, University of Victoria (2011)

Supervisory Committee

Dr. Stuart W. S. MacDonald, (Department of Psychology)
Co-Supervisor

Dr. Holly A. Tuokko, (Department of Psychology)
Co-Supervisor

Dr. Scott M. Hofer, (Department of Psychology)
Department Member

Dr. Sandra Hundza, (Department of Kinesiology and Exercise Science)
Outside Member

Abstract

Supervisory Committee

Dr. Stuart W. S. MacDonald, (Department of Psychology)
Co-Supervisor

Dr. Holly A. Tuokko, (Department of Psychology)
Co-Supervisor

Dr. Scott M. Hofer, (Department of Psychology)
Department Member

Dr. Sandra Hundza, (Department of Kinesiology and Exercise Science)
Outside Member

Objective: The purpose of this dissertation was to determine whether measures of reaction time inconsistency (RTI) can be applied clinically to detect cognitive impairment in older adults.

Methods: Data were obtained from the Victoria Longitudinal Study (VLS), a longitudinal study of healthy aging, and PREVENT, a multivariate study of risk factors for Alzheimer's disease. Study 1 examined effects of task complexity and computational approach on the association between RTI and physical and cognitive functioning in participants of the VLS. Study 2 assembled normative data from the VLS and standardized RTI data from an independent VLS cohort against these normative data. Significant Study 1 findings were replicated in Study 2 using the obtained RTI T-Scores, and the clinical utility of results were evaluated using stratum specific likelihood ratios (SSLRs). Study 3 replicated Study 2 analyses in data from PREVENT.

Results: Results of Study 1 identified four operationalizations of RTI from a choice reaction task that yielded consistent significant associations with cross-sectional cognitive performance. Consistent associations were not observed between these scores and cognitive change or performance on measures of physical functioning. Study 2 replicated Study 1 findings

in an independent sample using RTI T-Scores. SSLRs supported the clinical utility of measures of RTI for detecting prevalent cognitive impairment. Study 3 replicated findings from Study 2, but SSLRs indicated that only low RTI scores yielded associations of sufficient reliability for clinical interpretation. Consistent with Study 1 and Study 2, associations between RTI T-Scores and measures of physical function were nonsignificant.

Conclusions: Low RTI T-Scores were shown across two samples to be associated with a clinically meaningful reduction in the odds of cognitive impairment. Further research is needed in order to clarify the utility of high RTI scores for positive prediction of cognitive impairment.

Table of Contents

List of Tables

General Introduction

The study of brain-behaviour relationships has traditionally relied on measures of central tendency (Hultsch, Strauss, Hunter, & Macdonald, 2008). Implicit in this approach is the assumption that variance surrounding the mean is error and confers no additional meaningful information about group membership. There have been significant developments in statistical and methodological approaches that allow for departures from reliance on measures of central tendency, but adoption of these methods for the purposes of neuropsychological assessment has been slow. The reaction time (RT) literature has been at the forefront of approaches that accommodate violations of normality because the assumptions of central tendency are rarely met in these data. While classical test theory assumes that variability surrounding the true score is attributable to measurement error, reaction time inconsistency (RTI) has long been recognized as a source of meaningful information about psychological processes over and above what is conferred by the mean or median RT (Heathcote, Popiel, & Mewhort, 1991). There is now compelling evidence that measures of RTI are independent of mean RT, sensitive to the integrity of the central nervous system, and cross-sectionally and prospectively associated with cognitive outcomes in late life (e.g., Dykiert, Der, Starr & Deary, 2013; MacDonald, Nyberg, & Backman, 2006).

RTI in relation to central nervous system integrity

The current literature suggests that individuals with higher RTI have lower white matter volume (Anstey et a., 2007), more vascular lesions (Jackson, Balota, Duchek, & Head, 2012), decreased dopamine receptor binding (Macdonald, Cervenka, Farde, Nyberg, & Backman, 2009) and less distinct cortical representation of cognitive functions than those with lower RTI

(Macdonald, Nyberg, Sandblom, Fischer, & Backman, 2008). Significant findings have been observed between RTI and imaging markers of neural integrity both in clinical samples (Jackson, Balota, Duchek, & Head, 2012; Anstey et al., 2007; Stuss, Murphy, Binns, & Alexander, 2003; Murtha, Cismaru, Waechter, & Chertkow, 2002) and in individuals with no neuropsychological deficits (Walhovd & Fjell, 2007; Fjell et al., 2011; Moy et al, 2011; Bunce et al., 2007, Lovden et al., 2013; Macdonald, Nyberg, Sandblom, Fischer, & Backman, 2008). These findings have been shown to be independent of mean RT, and have been observed using measures of simple RT and more demanding choice and recognition RT tasks.

The imaging literature complements a larger body of behavioural research demonstrating significant positive associations between RTI and cognitive status. RTI is elevated in Alzheimer's disease (AD) relative to healthy controls (Hultsch, Macdonald, Hunter, Levy-Bencheton, & Strauss, 2000) and individuals with milder cognitive impairment (Gorus, De, Lambert, Lemper, & Mets, 2008). RTI is also elevated in individuals with mild cognitive impairment (MCI) relative to healthy older adults (Gorus, De, Lambert, Lemper, & Mets, 2008; Strauss, Bielak, Bunce, Hunter, & Hultsch, 2007; Dixon et al., 2007). The literature suggests that AD can be differentiated from normal aging using both simple and more complex RT tasks, but RT tasks of higher complexity are needed to elicit differences between individuals with MCI relative to healthy older adults (Gorus, De, Lambert, Lemper, & Mets, 2008). Higher RTI has further been shown to predict progression from normal aging to MCI over four years (Cherbuin, Sachdev, & Anstey, 2010), and from MCI to AD over three years (Tales et al., 2012). In addition, higher RTI has been observed in healthy individuals with the APOE E4 allele and cerebrospinal biomarkers for AD (Duchek et al., 2009), and in individuals with Type II diabetes (Whitehead, Dixon, Hultsch, & Macdonald, 2011). Null findings have also been reported in this

literature, such that RTI scores did not differentiate MCI from healthy aging independent of mean RT in one study (Christensen et al., 2005).

Clinical applications of RTI

On the basis of the reviewed literature implicating a unique association between RTI and cognitive dysfunction, many investigators have referenced the potential clinical utility of RTI for detecting impairment (Hultsch, Macdonald, Hunter, Levy-Bencheton, & Strauss, 2000; Lovden et al., 2013). However, the feasibility and validity of RTI for clinical use has not been examined. Steps and considerations involved in the development and validation of a clinical tool differ in critical ways from the development of an experimental measure. In particular, issues related to standardized assessment and norm-referenced testing, predictive validity, and evidence-based practice are all important considerations for the evaluation of a potential clinical measure of cognitive functioning, but are usually of no relevance or concern in the context of experimental research. The American Psychological Association's Standards for Psychological and Educational Testing (APA, 1999) provide guidelines for the development and evaluation of measures of psychological functioning, including guidelines for development of normative data and aspects of validity that should be examined in tests intended for clinical use. To date there have been no published attempts to develop or evaluate any measures of RTI for clinical use.

Standardized Assessment: Clinical neuropsychology, like other subdisciplines of clinical psychology, takes a standardized approach to assessment. Critical to the process of standardized assessment is normative comparison. Norm-referenced testing involves taking an observed test score and comparing it to the performance of a sample of individuals of a similar age as the examinee. Further stratification is carried out when demographic characteristics are found to contribute strongly to performance on a given test. Failure to appropriately stratify normative

data could result in bias against individuals with demographic backgrounds that are different from those of the normative sample.

Norm-referenced testing relies on the assumption that the abilities measured by a given test are normally distributed in the populations (Nunnally & Bernstein, 1994). Individuals who score more than one standard deviation above or below the mean are classified as having higher or lower aptitude for the abilities measured by a given test relative to others their age. Sufficiently extreme responses, usually classified as 1.5 to 2 standard deviations beyond the mean, are classified as impaired and reflective of a pathologically low aptitude for the abilities measured by a given test relative to examinees with similar characteristics.

The Standards for Psychological and Educational Testing (APA, 1999) outline criteria for developing normative data. These criteria emphasize the importance of ensuring that the normative data that is used to interpret an examinee's test score was obtained from a population that is truly demographically comparable to the examinee. Samples should also consist of at least 100 or more participants in total in order to ensure the reliability of normative estimates.

Criterion Validity: Measures of reaction time differ from more conventional cognitive measures in that their design typically includes a large number of trials of the same essential task (Nunnally & Bernstein, 1994). Thus, issues related to content validity, differential item functioning and internal consistency may be of less relevance. However, many issues in conventional psychological measurement do apply to measures of reaction time. For example, predictive validity, also referred to as criterion-related validity, refers to the value of a measure for predicting an independent variable or outcome. The Standards for Psychological and Educational Testing require that predictive validity of a measure be demonstrated before recommending that it be used with a new population (APA, 1999).

Evidence-Based Practice: The presence of a statistically significant association between a measure and an outcome provides little insight into the potential clinical utility of a measure (Akobeng, 2007). Thus, in order to effectively demonstrate validity, statistical approaches must be adopted that can provide meaningful insight into the predictive power of a test score. Evidence-based practice describes the use of statistical estimates of risks and benefits derived from empirical research on population samples in order to inform clinical decision making (Greenhalgh, 2010). Likelihood ratios are among the most popular statistics to inform clinical decision making. Likelihood ratios provide an estimate of the added value that a given test score will provide for prediction of an outcome over and above the pretest probability that the outcome is present in a given examinee. In other words, likelihood ratios provide evaluative information about the diagnostic utility of a test based on the modification the test result would make to the pretest probability of the presence of the outcome in a given examinee (Akobeng, 2007). Likelihood ratios are derived from sensitivity and specificity estimates associated with a test score by obtaining the ratio of the test's true positives relative to false positives. There are established guidelines that can then be followed to determine whether the test's contribution to prediction of an outcome is meaningfully different from the pretest probability.

Objectives

The overarching purpose of this dissertation is to examine the clinical utility of RTI for predicting physical and cognitive functioning in older adults. Data are recruited from the Victoria Longitudinal Study (VLS), a longitudinal study of healthy aging, and PREVENT, a multivariate study of risk factors for AD. Study 1 empirically tests the optimal operationalization of RTI for use in subsequent analyses, and characterizes raw associations between RTI scores and physical and cognitive outcomes in the VLS. Study 2 assembles demographically stratified normative data from the VLS and standardizes RTI data from an independent VLS cohort against these normative data. Significant Study 1 findings are subsequently replicated in Study 2 using the obtained RTI T-Scores, and the clinical utility of results are interpreted using stratum specific likelihood ratios. Finally, Study 3 replicates Study 2 analyses in data from PREVENT to determine the utility of norm-referenced RTI scores for detecting clinically significant cognitive and functional impairments in a sample with more rigorously characterized physical and cognitive function. Results of these studies are discussed in relation to feasibility issues of clinical applications of RTI and avenues for future research.

Study 1: Comparison of Prevalent Operationalizations of Reaction Time Inconsistency in Relation to Physical Function and Cognition in Participants of the Victoria Longitudinal Study

Reaction time inconsistency (RTI) is a property of reaction time data that has been shown to share significant associations with central nervous system (CNS) function in older adults. RTI is elevated in diseases of aging, including mild cognitive impairment and Alzheimer's disease (Gorus, De, Lambert, Lemper, & Mets, 2008), cerebrovascular disease (Bunce et al., 2007), and Parkinson's disease (de Frias, Dixon, Fisher, & Camicioli, 2007). RTI is also elevated in diseases known to affect CNS functioning, such as diabetes (Whitehead, Dixon, Hultsch, & Macdonald, 2011). Higher RTI yields cross-sectional (e.g., Hultsch, Macdonald, Hunter, Levy-Bencheton, & Strauss, 2000) and prospective (e.g., Tales et al., 2012) associations with cognitive impairment and cognitive decline (Bielak, Hultsch, Strauss, Macdonald, & Hunter, 2010) in older adults. RTI is also associated with poorer performance on indicators of physical vitality, including grip strength, peak expiratory flow (Anstey, Dear, Christensen, & Jorm, 2005), and incident mortality (Macdonald, Hultsch, & Dixon, 2008). RTI associated with measures of both simple and complex RT have been shown to predict neural (e.g., Fjell, Westlye, Amlien, & Walhovd, 2011) and behavioural (e.g., MacDonald, Hultsch and Dixon, 2003) integrity, with a positive association between the degree of executive control involved in a task and the strength of its associated RTI scores and cognitive outcomes (e.g., West, Murphy, Armilio, Craik, & Stuss, 2002; (Gorus, De, Lambert, Lemper, & Mets, 2008). This behavioural literature is complemented by an emerging body of evidence demonstrating sensitivity of RTI obtained from tests of simple and complex RT to imaging markers of neural integrity (Lovden et al., 2013), and biological markers of neuropathology (Duchek et al., 2009). It has been suggested on the basis of

this research that RTI may represent a promising indicator of CNS integrity for use in clinical settings (Duchek et al., 2009, Hultsch et al.). However, before a credible attempt can be made to extend RTI into the clinical realm, the heterogeneity in the literature regarding the most conceptually and empirically defensible approach for operationalizing RTI must be addressed.

Many approaches to the estimation of RTI have been implemented to date, ranging from gross estimates based on the raw intraindividual standard deviation (ISD) to distribution-based based parameters obtained from mathematical models (e.g., Jackson, Balota, Duchek, & Head, 2012). For example, RTI has been operationalized using the raw and mean-partialed ISD, ISD estimates obtained from fast or slow tails of response distributions (e.g., Hultsch, Macdonald, & Dixon, 2002), the coefficient of variation (CoV; Jackson, Balota, Duchek, & Head, 2012), mean-absolute residuals (Anstey et al., 2007), the interquartile range (Dykiert, Der, Starr, & Deary, 2012), Ratcliff, shifted-Wald and Ex-Gaussian parameters (Matzke & Wagenmakers, 2009), and the mean square successive difference of RT trials (Santhanam, Simon, Seaman, Howard & Howard, 2013). Significant associations have been obtained using all of these metrics implicating higher RTI as indicative of the presence of CNS dysfunction. However, the strength of the associations among these RTI computations, and the extent to which they reflect the same central construct has yet to be demonstrated in relation to cognitive outcomes.

The most widely used operationalization of RTI in the literature is the ISD. Raw ISDs have been shown to be sensitive to age effects such that older adults yield higher ISDs than younger people (Dykiert, Der, Starr, & Deary, 2012). Raw ISDs are also elevated in individuals with cognitive impairment and dementia independent of mean reaction time. However, the use of raw ISDs has been criticised on conceptual grounds: it has been suggested that group comparison of ISD scores may yield biased findings if the groups are known to differ in mean reaction time.

In addition, raw ISDs capture learning effects across trials that may obfuscate the association between "pure" RTI and CNS outcomes (MacDonald, Hultsch and Dixon, 2003). To address this, many investigators have partialed between person differences in these characteristics from RT data prior to calculation of ISD scores. Early work implemented an ANCOVA-based approach to partialing of RT data (e.g., Hultsch, Macdonald, Hunter, Levy-Bencheton, & Strauss, 2000), but regression-based approaches are now considered preferable because they allow for partialing at the individual level rather than the group level. Partialled ISDs have been found to increase as a function of age and cognitive status at a magnitude similar to raw ISDs (Dykiert, Der, Starr, & Deary, 2012).

A different approach to operationalizing RTI using ISDs has involved restricting calculation of variability to fast and slow tails of intraindividual RT distributions. This approach follows the hypothesis that attentional/executive lapses drive the association between RTI and CNS outcomes, which lead to increases in the "slow" tail of responses in the intraindividual distribution. Results using percentile-based ISDs have generally supported this finding, with variability in the slow tail of the RT distribution correlating more strongly with ISDs computed from the whole distribution and sharing the stronger association with cognitive outcomes (Hultsch, Macdonald, & Dixon, 2002). Calculation of RTI based on fast vs. slow tails of the RT distribution can be thought of as a rudimentary approximation of the Ex-Gaussian "tau" parameter. While conceptually appealing and less computationally intensive than Ex-Gaussian parameters, research examining RTI from fast vs. slow tails of the RT distribution is limited.

The coefficient of variation (CoV), which represents the ratio of the standard deviation to the mean RT, is another common method of operationalizing RTI in aging research. The CoV is conceptually appealing because it inherently adjusts for the mean RT associated with a given

level of variability. Research examining the CoV has found it to be sensitive to Alzheimer's disease (Murtha, Cismaru, Waechter, & Chertkow, 2002) and white matter integrity (Jackson, Balota, Duchek, & Head, 2012; Bielak et al., 2013). It has further been found that CoV shares associations with CNS outcomes that are comparable to those observed for ex-Gaussian parameters (Jackson, Balota, Duchek, & Head, 2012), and the ISD (Batterham, Bunce, Mackinnon & Christensen, 2014). However, research examining age effects of RTI found the CoV to yield associations with age that were smaller in magnitude than the partialed ISD (Dykiert, Der, Starr, & Deary, 2012). Thus, while the CoV has some conceptually appealing properties, more research is needed in order to clarify its sensitivity to cognitive outcomes relative to other operationalizations of RTI.

Still another approach for computing RTI, and the method that has received the least attention in behavioural aging research, is the mean square successive difference of RT trials (MSSD). This approach involves computing the intraindividual mean of the sum of the squared difference between adjacent trials. As a result, the only difference between the MSSD and conventional estimates of variance is the fact that each RT value is compared to the immediately preceding RT value, rather than the overall RT mean (Garrett, Samanez-Larkin, MacDonald, Lindenberger, McIntosh & Grady, 2013). The MSSD as an estimate of variability is thus less affected by gradual shifts in RT values, such as potential practice and fatigue effects, and more sensitive to larger discrepancies in trial-to-trial RT. Although it is used widely to quantify variability in biometric data (e.g., heart rate, blood pressues), to date, only one study has applied the MSSD to the study of cognitive aging (Santhanam, Simon, Seaman, Howard & Howard, 2013). Results of this study were consistent with those reported using other operationalizations of RTI.

As reviewed here, many intriguing approaches to operationalizing RTI have been reported in the literature, but evidence for their relative sensitivity to CNS outcomes, especially cognitive outcomes, is lacking. The purpose of the current study is to systematically evaluate the relative association of eight operationalizations of RTI to baseline and longitudinal performance on tests of physical and cognitive function. In addition, this study examines the relative sensitivity of these RTI operationalizations to cognitive status in the same sample. Results of this study will inform subsequent work addressing the potential clinical utility of RTI for detecting cognitive impairment. These research questions are addressed using data from both simple and choice RT tasks in order to determine the importance of task complexity for yielding RTI scores that are sensitive to individual differences in physical and cognitive function.

Study 1 Methods

The Victoria Longitudinal Study (VLS) is a longitudinal study of multiple facets of human aging. The design of the VLS is described in detail elsewhere (Dixon & de Frias, 2004; Hultsch, Hertzog, Dixon & Small, 1998). Following a longitudinal sequential research design, the VLS includes three cohorts of participants, all aged 55-85 at their baseline assessment, which undergo testing in 3-year intervals. Sample 1 began in 1986 with 484 participants, Sample 2 included 530 participants who were first assessed in 1992, and Sample 3 consists of 550 participants who were first tested in 2001. To date, Sample 1 has completed 7 assessments over 18 years, Sample 2 has completed 5 assessments over 12 years, and Sample 3 has completed 3 assessments over 6 years. Participants of the VLS were recruited from the community and were free of serious health conditions at study entry. Specific exclusionary criteria at baseline included a diagnosis of Alzheimer's disease or any other neurological disorder, presence of any psychiatric conditions or medications, preexisting serious cardiovascular or cerebrovascular conditions, corrected eyesight insufficient for reading, and corrected hearing insufficient for comprehension of spoken instructions (Dixon & de Frias, 2004). This study examined baseline and longitudinal data obtained from Sample 3.

Cognitive Measures

The following cognitive measures were included in the core battery that was administered routinely to participants across waves of the VLS:

Letter Series Task: Reasoning was assessed using the letter series task (Thurstone, 1962). Participants were presented with strings of letters that followed a particular pattern and tasked with providing the next letter that followed the pattern. Participants were given six minutes to complete 20 strings of letters.

Word List Recall: Episodic memory was assessed through participant's immediate recall of 30 English words, each falling within one of five semantic categories. Participants were given two minutes to study the list and five minutes to write down all the words that they could recall.

Vocabulary: Vocabulary was assessed with a 36 item multiple choice test where participants were given ten minutes to select the correct definition of each word from five possible definitions (Ekstrom, French, Harman, & Dermen, 1976). The timed aspect of this task distinguishes it from conventional measures of vocabulary.

Similarities: Verbal abstraction was assessed using the Similarities subtest of the WAIS-R (Wechsler, 1981). This task requires participants to identify commonalities among objects or concepts.

Digit Symbol: Perceptual Speed was assessed using the Digit-Symbol Substitution subtest of the WAIS-R (Wechsler, 1981). This task requires participants to follow a coding key and assign rows of numbers with their corresponding shape as indicated in the coding key. Participants had up to 90 seconds to complete as many items as possible.

Measures of Physical Functioning

Measures of biological vitality collected in the VLS include blood pressure, peak expiratory flow, grip strength, body mass index, balance, gait, and self-report information about physical health and medical conditions. The present study examined associations between RTI and objective measures of systolic blood pressure, diastolic blood pressure, pulse, peak flow and grip strength. These variables were examined because of their well-documented association with physical and cognitive vitality in the elderly (e.g., DeCarlo, Tuokko, Williams, Dixon & MacDonald).

Reaction Time Measures

Starting at Wave 3 of Sample 1, RT data were obtained at each testing occasion. A measure of simple RT (SRT) was included along with 2, 4 and 6-choice reaction time tasks and tasks requiring participants to distinguish words from non-words (lexical decision task), and plausible sentences from implausible sentences (semantic verification task). The SRT and Lexical Decision tasks have been most thoroughly investigated in relation to RTI, and are described in more detail here:

Simple Reaction Time (SRT): The SRT is a computerized measure that presents participants with a warning stimulus (asterisks) followed by a signal stimulus (plus sign) in the middle of the computer screen. Participants are tasked with pressing a key as quickly as possible following the appearance of the signal stimulus. The VLS included 50 test trials of the SRT, with five inter-stimulus intervals (500, 625, 750, 875, and 1,000 ms) distributed evenly across trials (e.g., each inter-stimulus interval assigned to 10 trials). Trials were presented to participants in random order and latencies of the 50 trials form the outcome measure of this task.

Lexical Decision Making: The Lexical Decision task involves making rapid judgments regarding whether a string of 5-7 letters, as presented, formed an English word (e.g., salad vs. neefle). Participants are tasked with pressing one of two keys, depending on their response (e.g., press button one if the letters form a word, press button two if the letters do not form a word). Participants of the VLS completed 60 randomly ordered test trials (30 words, 30 non-words) at each measurement occasion. Latencies of these 60 trials form the task's outcome measure.

Operationalization of RTI

The raw ISD was obtained by calculating the standard deviation of each Sample 3 participant's performance across each RT task at Wave 1. Calculation of residual RTI scores

involved regressing trial-level RT data on age, sex, trial and the interaction among these variables using linear regression. Residual ISDs were then computed from the residualized RT data. The CoV was calculated by dividing the raw ISD of each participant by their raw intraindividual mean. The MSSD was obtained by computing the intraindividual mean of the sum of the squared difference between adjacent trials.

Fast and slow ISDs were obtained by ranking each participant's RT data by latency. For example, each participant's RT data for each of the 50 trials or the SRT were ranked from fastest to slowest and converted to percentiles. For each examinee, RT trials ranking within their fastest 20% of responses were used to calculate their "fast" ISD and trials ranking within the slowest 20% of responses were used to calculate their "slow" ISD. Residual fast and slow ISDs were computed by applying the same procedures to residualized RT data.

Operationalization of Cognitive Functioning

RTI was examined in relation to cognitive performance in Sample 3 using both continuous and discrete operationalizations of cognitive functioning. Continuous operationalizations were raw scores associated with the letter series, word list recall, vocabulary, digit symbol and verbal fluency measures. RTI scores were examined in relation to baseline cognitive performance (e.g., performance of Sample 3 at Wave 1) and in relation to longitudinal change in cognitive performance across three measurement occasions and five years. RTI scores were subsequently examined in relation to several operationalizations of cognitive status.

Several investigators have adopted a distributional approach to operationalizing cognitive impairment in the VLS by classifying individuals falling 1 SD below the sample mean on any cognitive test as having mild cognitive impairment (MCI; e.g., Strauss, Bielak, Bunce, Hunter, & Hultsch, 2007). Vandermorris and colleagues (2011) subsequently demonstrated that

participants meeting study criteria for MCI across two or more consecutive study assessments had lower baseline cognition and steeper cognitive decline than individuals with no cognitive impairment or those meeting criteria for MCI at only one measurement occasion (Vandermorris, Hultsch, Hunter, MacDonald & Strauss, 2011). For the purposes of examining the sensitivity of RTI to cognitive functioning in the VLS, the multiple-assessment MCI (MA-MCI) classification developed by Vandermorris and colleagues (2011) served as the primary operationalization of cognitive status in this study. However, for exploratory purposes, several other operationalizations of cognitive status, computed to correspond with varying levels of impairment severity, were also examined.

To examine the sensitivity of RTI to the mildest of memory deficits, we classified participants with 1) memory performance falling 1 SD or more below the mean for their level of age and education, and 2) no other test scores falling more than 0.5 SD below the mean, as mild single-domain amnestic MCI. We subsequently classified those participants with 1) memory performance falling 1.5 SD below the mean for their age and level of education and 2) performance on at least one other cognitive test falling at least 1 SD below the mean as multidomain MCI. Participants with 1) memory performance falling 2 SD below the mean for their age and level of education and 2) performance on at least one other cognitive test falling at least 1.5 SD below the mean were classified as moderate multidomain MCI. Application of these criteria to data from Sample 3, Wave 1 participants of the VLS identified 57 participants (10%) meeting criteria for mild single-domain amnestic MCI, 46 participants (8%) meeting criteria for mild multidomain MCI, 29 participants (5%) meeting criteria for moderate multidomain MCI, and 69 participants (12%) meeting criteria for MA-MCI.

Operationalization of Physical Functioning

Functional status was operationalized using scores associated with objective measures of systolic blood pressure, diastolic blood pressure and pulse (average recording across 2 measurements), grip strength (average recording across 2 measurements), and peak flow (average recording across 2 measurements). This study examined baseline measurements (e.g., Sample 3 Wave 1 performance), and longitudinal change in objective measurements across three measurement occasions over five years.

Statistical Analyses

Data Preparation. Trial-level data from the SRT and Lexical Decision tasks were first screened for outliers. Following prior research, reaction times of 150 ms or less for the SRT task, and those 400 ms or less for the Lexical Decision task were not included in computations of any operationalization of RTI. In addition, reaction times falling three or more standard deviations above each participant's intraindividual mean were also excluded from analysis. These steps were taken to optimize comparability of our findings with prior research, and to ensure that characteristics of the reaction time distribution were not influenced by external sources of measurement error, such as accidental button presses and distraction of the participant. Following removal of trial-level outliers, computation of RTI scores proceeded as described previously (pg. 18).

Univariate and Bivariate Analyses: We examined the skewness and kurtosis values associated with the computed RTI values prior to examination of their association with study outcomes. Reaction time data, including RTI, is notoriously skewed (e.g., Heathcote, Popiel, & Mewhort, 1991). Following the approach adopted by developers of the Conners' CPT II (Conners & MHS Staff, 2000), log-10 transformation was applied to RTI data. Log-10

transformations convert data points into the power of 10 needed to obtain the observed data point. For example, a log-10 transformation would convert a value of 1000 into 3, because 1000 is equal to 10^3. Log-transformed data naturally conform to a normal distribution, thus making it more compatible with the assumptions of normality that are inherent in regression-based analytical approaches. Associations among the eight RTI variables and associations between log-transformed and untransformed RTI scores were subsequently examined to determine the comparability of these values.

Linear Regression Analyses: Linear regression was used to examine the association between RTI scores and baseline performance on the physical and cognitive study outcomes. All linear regression analyses were adjusted for age, sex, education (measured as a continuous variable), and mean reaction time on the SRT (for analyses examining SRT RTI scores) or the Lexical Decision task (for analyses examining Lexical RTI scores). Analyses were replicated using log-transformed RTI scores to determine whether associations observed from unadjusted RTI scores were influenced by deviations from normality. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Mixed Linear Modeling: Within-person change in cognitive scores and performance on measures of physical functioning over the three waves of Sample 3 of the VLS was estimated in relation to baseline RTI performance using mixed linear modeling (MLM). Mixed linear modeling allows for the assessment of within-person change over time (Level 1), and between-person differences in within-person change (Level 2). All mixed models were adjusted for mean RT at baseline, age, sex and education. Models were further adjusted for random effects associated with age at baseline. Time in study was selected as the metric for time (Level 1) because it provides the best parameterization of time and circumvents age convergence issues

associated with the use of age as time (Morrell, Brant, & Ferrucci, 2009). Full information maximum likelihood was used for parameter estimation. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Logistic Regression Analyses: Logistic regression was used to examine the association between baseline SRT and Lexical RTI scores and each of the four cognitive outcomes. Logistic regression models the association between a binary outcome and one or more predictors. The presence (coded as "1") or absence (coded as "0") of cognitive impairment will form the outcome. RTI scores were examined as predictors of cognitive status along with age, sex, education, and mean RT. ROC curves associated with RTI scores were examined in order to determine the potential classification accuracy of these values and the likelihood ratio test was used to determine the contribution of RTI scores to prediction of cognitive status over and above mean RT and model covariates. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Study 1 Results

Participants

Participant characteristics are presented by cognitive status in Table 1. Participants classified as MA-MCI had approximately one less year of education and slower mean RT values than the healthy group. The sample was very young, with a mean age under 70 in both the healthy and impaired groups. Groups did not differ in performance on measures of physical function, in their average age, or in the gender distribution of the group.

Table 1. Demographic characteristics of the VLS sample.

	Healthy (n=482)	MA-MCI (n=67)	*F*(*df*), *p*-value
Age	68.13 (8.71)	69.13 (7.78)	0.331 (1) *NS*
Sex, %F	69%	58%	3.45 (1) *NS*
Education	**15.31 (2.94)**	**14.52 (3.52)**	**4.39 (1)***
Systolic BP	126.08 (15.37)	128.08 (22.12)	0.900 (1) *NS*
Grip Strength	30.99 (9.69)	33.15 (9.54)	0.699 (1) *NS*
Peak Flow	419.22 (115.48)	437.81 (120.99)	1.778 (1) *NS*
Mean Lexical	**1067.10 (398.71)**	**1260.52 (455.11)**	**13.34 (1)****

Note. VLS = Victoria Longitudinal Study; MA-MCI = multi-assessment mild cognitive impairment; NS = nonsignificant; BP = blood pressure; *df* = degrees of frequency. All values are presented as mean (standard deviation) unless noted otherwise.

Table 2a. Distributional characteristics of raw and log-transformed SRT RTI scores in Sample 3 of the VLS

	Raw RTI Scores						Log-Transformed RTI Scores					
RTI Scores for SRT	RTI Mean	SD	Skew	SE Skew	Kurt	SE Kurt	RTI Mean	SD	Skew	SE Skew	Kurt	SE Kurt
Raw ISD	87.79	58.904	3.43	0.103	16.191	0.206	1.884	0.210	0.836	0.103	1.554	0.206
Res ISD	0.693	0.465	3.44	0.103	16.300	0.206	-0.219	0.210	0.840	0.103	1.578	0.206
CoV	0.266	0.150	3.04	0.103	11.884	0.206	-0.620	0.183	0.956	0.103	1.883	0.206
MSSD	118.45	81.562	3.482	0.103	16.299	0.206	2.013	0.211	0.924	0.103	1.696	0.206
Raw Fast ISD	15.137	9.720	2.193	0.103	6.826	0.206	1.111	0.240	0.267	0.103	0.171	0.206
Res Fast ISD	0.289	0.185	2.194	0.103	6.835	0.206	-0.608	0.238	0.305	0.103	0.104	0.206
Raw Slow ISD	88.150	122.22	4.080	0.103	21.017	0.206	1.781	0.339	0.819	0.103	1.137	0.206
Res Slow ISD	0.484	0.611	4.097	0.103	21.157	0.206	-0.475	0.329	0.951	0.103	1.198	0.206

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SD = standard deviation; SRT = simple reaction time; SE = standard error.

Table 2b. Distributional characteristics of raw and log-transformed Lexical Decision Task RTI scores in Sample 3 of the VLS

	Raw RTI Scores						Log-Transformed RTI Scores					
Lexical RTI Scores	RTI Mean	SD	Skew	SE Skew	Kurt	SE Kurt	RTI Mean	SD	Skew	SE Skew	Kurt	SE Kurt
Raw ISD	344.199	240.721	3.101	0.103	15.133	0.206	2.465	0.236	0.573	0.103	0.346	0.206
Res ISD	0.602	0.422	3.162	0.103	15.496	0.206	-0.291	0.234	0.631	0.103	0.455	0.206
CoV	0.296	0.098	1.040	0.103	1.646	0.206	-0.551	0.138	0.152	0.103	-0.277	0.206
MSSD	428.444	306.353	3.174	0.103	15.349	0.206	2.559	0.237	0.628	0.103	0.468	0.206
Raw Fast ISD	51.869	34.796	5.615	0.103	60.273	0.206	1.657	0.214	0.454	0.103	1.201	0.206
Res Fast ISD	0.106	0.064	5.469	0.103	56.211	0.206	-1.023	0.192	0.620	0.103	1.611	0.206
Raw Slow ISD	210.859	152.602	2.923	0.103	17.529	0.206	2.238	0.270	0.163	0.103	-0.069	0.206
Res Slow ISD	0.390	0.289	2.973	0.103	18.100	0.206	-0.500	0.278	0.113	0.103	-0.007	0.206

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SD = standard deviation; SE = standard error.

Univariate Characteristics of RTI Scores

RTI scores were calculated as described in the Methods. Distributional characteristics of the obtained RTI scores are presented in Table 2a and 2b. All RTI scores were positively skewed and several operationalizations exhibited significant kurtosis. Normality of the SRT and Lexical Decision RTI distributions was similar (SRT skewness range: 2.19-4.10, kurtosis range: 6.83-21.16; Lexical skewness range: 1.04-5.61; kurtosis range: 1.65-60.27). However, for the SRT there tended to be greater kurtosis in ISDs obtained from the slowest 20% of responses relative to the fastest 20% of responses (e.g., kurtosis values of 21.157 vs. 6.835). In contrast, the Lexical Decision task was associated with much higher kurtosis values for ISD scores obtained from the fastest 20% of responses relative to the slowest (e.g., kurtosis values (e.g., 56.211 vs. 18.100). This observation may suggest that variability in this sample is more heterogeneous for slow responses on the SRT relative to fast responses, and for fast responses on the Lexical Decision task relative to slow responses. Due to clear violations of normality for RTI scores obtained from both RT tasks, subsequent analyses were conducted both on raw and log-10-transformed RTI scores in order to determine whether non-normality of scores influenced their association with study outcomes. Distributional characteristics of Log-transformed RTI scores approximated normality, and are also presented in Table 2a and 2b.

Bivariate Associations among RTI Scores

Bivariate associations among mean RT and the eight operationalizations of RTI for each task are presented in Table 3a and 3b. For the SRT, the raw and residual ISD yielded near-identical associations with mean RT (0.434 vs. 0.433), and the MSSD yielded an association of a similar magnitude with mean RT (0.380). The strongest association with mean RT was observed for ISDs obtained from the fastest 20% of responses (0.689). In contrast, the CoV and ISD scores

Table 3a. Correlation matrices for raw and log-transformed SRT RTI Scores.

	Raw RTI Scores								Log-Transformed RTI Scores							
	Raw ISD	Res ISD	CoV	MSSD	Raw fast ISD	Res fast ISD	Raw slow ISD	Res slow ISD	Raw ISD	Res ISD	CoV	MSSD	Raw fast ISD	Res fast ISD	Raw slow ISD	Res slow ISD
SRT Mean	**.43**	**.43**	**.09**	**.38**	**.69**	**.69**	**.17**	**.17**	**.49**	**.49**	**.08**	**.44**	**.67**	**.67**	**.17**	**.17**
Raw ISD	**1.00**	**.99**	**.92**	**.98**	**.24**	**.24**	**.91**	**.91**	**1.00**	**.99**	**.91**	**.97**	**.29**	**.29**	**.82**	**.82**
Res ISD		**1.00**	**.92**	**.98**	**.24**	**.24**	**.91**	**.91**		**1.00**	**.91**	**.97**	**.29**	**.29**	**.82**	**.82**
CoV			**1.00**	**.92**	-.01	-.01	**.93**	**.93**			**1.00**	**.89**	.00	-.00	**.87**	**.86**
MSSD				**1.00**	**.20**	**.20**	**.91**	**.91**				**1.00**	**.27**	**.27**	**.81**	**.81**
Raw fast ISD					**1.00**	**.99**	.04	.04					**1.00**	**.99**	.04	.04
Res fast ISD						**1.00**	.04	.04						**1.00**	.04	.03
Raw slow ISD							**1.00**	**.99**							**1.00**	**.99**
Res slow ISD								**1.00**								**1.00**

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time. Bold values denote significant associations.

Table 3b. Correlation matrices for raw and log-transformed Lexical Decision Task RTI Scores.

	Raw RTI Scores								Log-Transformed RTI Scores							
	Raw ISD	Res ISD	CoV	MSSD	Raw fast ISD	Res fast ISD	Raw slow ISD	Res slow ISD	Raw ISD	Res ISD	CoV	MSSD	Raw fast ISD	Res fast ISD	Raw slow ISD	Res slow ISD
Lexical Mean	**.89**	**.89**	**.54**	**.88**	**.83**	**.78**	**.70**	**.70**	**.81**	**.82**	**.51**	**.81**	**.74**	**.70**	**.66**	**.66***
Raw ISD	**1.00**	**.99**	**.82**	**.96**	**.70**	**.66**	**.88**	**.88**	**1.00**	**.99**	**.90**	**.96**	**.66**	**.62**	**.90**	**.89**
Res ISD		**1.00**	**.82**	**.97**	**.70**	**.66**	**.88**	**.88**		**1.00**	**.89**	**.96**	**.66**	**.62**	**.89**	**.89**
CoV			**1.00**	**.76**	**.36**	**.34**	**.84**	**.84**			**1.00**	**.83**	**.41**	**.38**	**.85**	**.84**
MSSD				**1.00**	**.70**	**.66**	**.84**	**.84**				**1.00**	**.66**	**.63**	**.87**	**.87**
Raw fast ISD					**1.00**	**.93**	**.48**	**.48**					**1.00**	**.84**	**.54**	**.54**
Res fast ISD						**1.00**	**.45**	**.45**						**1.00**	**.51**	**.50**
Raw slow ISD							**1.00**	**.99**							**1.00**	**.97**
Res slow ISD								**1.00**								**1.00**

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

obtained from the slowest 20% of responses yielded small associations with mean RT (0.09, 0.171). Correlations among SRT RTI scores revealed strong associations between raw and residual ISD and MSSD, CoV, and ISD scores obtained from the slowest 20% of responses. ISD scores obtained from the fastest 20% of responses did not correlate highly with any of the other RTI scores obtained from the SRT (-0.014-0.235). Correlations among log-transformed SRT RTI scores, presented in Table 3a, demonstrated a very similar pattern of associations. Correlations between log-transformed and untransformed SRT RTI scores are presented in Table 4.

Table 4. Correlations between log-transformed and untransformed RTI scores.

RTI Score	SRT	Lexical
Raw ISD	**.912****	**.911****
Res ISD	**.912****	**.911****
CoV	**.935****	**.978****
MSSD	**.913****	**.909****
Raw fast ISD	**.924****	**.866****
Res fast ISD	**.926****	**.881****
raw slow ISD	**.843****	**.898****
Res slow ISD	**.851****	**.892****

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

For the Lexical Decision task, the CoV yielded a moderate association with mean RT (0.501) and all other RTI scores yielded strong associations with mean RT (0.700-0.891). Correlations among lexical RTI scores revealed strong associations among most of the 8 examined scores. Relative to the other RTI scores, ISD scores obtained from the fastest 20% of responses yielded the weakest associations with other scores (0.363-0.704). Correlations among log-transformed Lexical RTI scores, presented in Table 3b, yielded a very similar pattern of associations. Associations between log-transformed and untransformed Lexical RTI scores, presented in Table 4, were strong (0.866-0.978).

RTI Scores in Relation to Physical Functioning at Baseline

SRT: The association between RTI scores for the SRT and objective measures of baseline grip strength, peak flow, pulse, systolic blood pressure and diastolic blood pressure was examined using a series of linear regression models adjusted for age, sex, education and mean reaction time on the SRT. Results are presented in Table 5a. No significant association was observed between any of the SRT RTI scores and any of the five measures of physical functioning. Null findings were similarly observed for log-transformed RTI scores.

Lexical Decision: Linear regression models adjusted for age, sex, education and mean reaction time on the Lexical Decision task were used to examine the relationship between RTI scores and objective measures of baseline physical functioning. Pulse was significantly associated with the raw MSSD, but this association fell below our criterion for significance for the log-transformed MSSD. No other significant observations were observed between Lexical RTI scores and baseline measures of physical functioning.

Table 5a. Linear regression of SRT RTI scores on objective measures of physical function.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	R2 Adj	B (95% CI)	*P*	R2 Adj
Systolic BP: Ref R2			.091			.091
Raw ISD	-0.008 (-0.033, 0.017)	.510	.090	-3.997 (-11.245, 3.250)	.279	.092
Res ISD	-1.059 (-4.213, 2.094)	.510	.090	-3.974 (-11.227, 3.279)	.282	.092
CoV	-4.247 (-13.146, 4.652)	.349	.091	-3.909 (-11.209, 3.391)	.293	.092
MSSD	-0.007 (-0.025, 0.010)	.422	.091	-2.989 (-10.034, 4.057)	.405	.091
Raw Fast ISD	-0.015 (-0.203, 0.172)	.871	.090	-1.309 (-8.765, 6.146)	.730	.090
Res Fast ISD	-0.961 (-10.817, 8.894)	.848	.090	-1.577 (-9.088, 5.934)	.680	.090
Raw Slow ISD	-0.005 (-0.016, 0.007)	.452	.091	-1.649 (-5.611, 2.314)	.414	.091
Res Slow ISD	-0.827 (-3.016, 1.362)	.458	.091	-1.564 (-5.632, 2.504)	.450	.091
Diastolic BP: Ref R2			.025			.025
Raw ISD	-0.003 (-0.018, 0.012)	.693	.023	-0.868 (-5.131, 3.395)	.689	.023
Res ISD	-0.384 (-2.238, 1.470)	.684	.023	-0.889 (-5.155, 3.378)	.683	.023
CoV	-1.275 (-6.508, 3.957)	.632	.023	-0.684 (-4.978, 3.610)	.754	.023
MSSD	-0.002 (-0.013, 0.008)	.653	.023	-0.257 (-4.400, 3.895)	.903	.023
Raw Fast	-0.002 (-0.112, 0.109)	.976	.023	-1.283 (-5.663, 3.098)	.565	.023

ISD						
Res Fast ISD	-0.156 (-5.947, 5.636)	.958	.023	-1.497 (-5.892, 2.933)	.510	.024
Raw Slow						
ISD	-0.002 (-0.009, 0.005)	.537	.024	-0.131 (-2.460, 2.199)	.912	.023
Res Slow						
ISD	-0.394 (-1.681, 0.892)	.547	.023	0.041 (-2.351, 2.433)	.973	.023
Pulse: Ref R2			.008			.008
Raw ISD	0.006 (-0.008, 0.021)	.383	.008	1.368 (-2.834, 5.570)	.523	.007
Res ISD	0.843 (-0.983, 2.670)	.365	.008	1.533 (-2.672, 5.739)	.474	.007
CoV	2.916 (-2.238, 8.071)	.267	.009	1.546 (-2.686, 5.778)	.473	.007
MSSD	0.004 (-0.006, 0.015)	.395	.008	1.298 (-2.786, 5.381)	.588	.007
Raw Fast						
ISD	0.065 (-0.044, 0.173)	.242	.009	-0.132 (-4.452, 4.189)	.952	.006
Res Fast ISD	3.259 (-2.445, 8.963)	.262	.009	-0.268 (-4.621, 4.084)	.904	.006
Raw Slow						
ISD	0.002 (-0.004, 0.009)	.491	.007	0.063 (-2.235, 2.360)	.957	.006
Res Slow						
ISD	0.422 (-0.846, 1.690)	.514	.007	0.058 (-2.301, 2.416)	.962	.006
Peak Flow: Ref R2			.453			.453
Raw ISD	-0.007 (-0.154, 0.140)	.924	.452	-3.334 (-44.16, 37.50)	.873	.452
Res ISD	-0.733 (-19.36, 17.90)	.938	.452	-2.631 (-43.50, 38.24)	.899	.452
CoV	4.933 (-46.18, 56.15)	.850	.452	-6.030 (-47.19, 35.13)	.774	.453
MSSD	-0.013 (-0.117, 0.091)	.808	.452	-8.703 (-48.49, 31.08)	.668	.453

Raw Fast						
ISD	-0.240 (-1.288, 0.809)	.854	.453	-13.967 (-55.20, 27.26)	.506	.453
Res Fast ISD	-12.218 (-67.27, 42.83)	.663	.453	-12.998 (-54.55, 28.55)	.539	.453
Raw Slow						
ISD	0.021 (-0.050, 0.093)	.562	.453	-7.015 (-35.46, 21.431)	.628	.451
Res Slow						
ISD	3.951 (-9.168, 17.07)	.554	.453	-7.614 (-35.16,19.93)	.587	.451
Grip Strength: Ref R2			.658			.658
Raw ISD	0.001 (-0.008, 0.010)	.785	.657	0.242 (-2.376, 2.860)	.856	.657
Res ISD	0.169 (-0.970, 1.307)	.771	.657	0.317 (-2.304, 2.937)	.812	.657
CoV	0.773 (-2.422, 3.969)	.635	.657	0.421 (-2.218, 3.059)	.754	.657
MSSD	0.001 (-0.005, 0.007)	.763	.657	0.318 (-2.215, 2.851)	.805	.657
Raw Fast						
ISD	0.023 (-0.047, 0.093)	.524	.657	0.690 (-2.029, 3.409)	.618	.657
Res Fast ISD	1.154 (-2.516, 4.824)	.537	.657	0.638 (-2.103, 3.378)	.648	.657
Raw Slow						
ISD	0.001 (-0.003, 0.005)	.642	.657	0.351 (-1.101, 1.803)	.635	.657
Res Slow						
ISD	0.185 (-0.608, 0.979)	.646	.657	0.305 (-1.177, 1.788)	.686	.657

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time; BP = blood pressure

Table 5b. Linear regression of Lexical RTI scores on objective measures of physical function.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	R2 Adj	B (95% CI)	*P*	R2 Adj
Systolic BP: Ref R2			.092			.092
Raw ISD	-0.005 (-0.017, 0.007)	.444	.092	-4.521 (-14.162, 5.121)	.357	.092
Res ISD	-3.473 (-10.349, 3.402)	.321	.092	-6.599 (-16.433, 3.235)	.188	.093
CoV	-7.960 (-24.027, 8.106)	.331	.092	-5.099 (-16.270, 6.073)	.370	.092
MSSD	0.000 (-0.009, 0.009)	.970	.091	-1.908 (-11.416, 7.600)	.694	.091
Raw Fast ISD	-0.012 (-0.080, 0.055)	.719	.091	-0.691 (-9.712, 8.331)	.880	.091
Res Fast ISD	-6.515 (-39.92, 26.89)	.702	.091	0.627 (-8.980, 10.234)	.898	.091
Raw Slow ISD	-0.011 (-0.023, 0.001)	.084	.096	-5.031 (-11.597, 1.536)	.133	.094
Res Slow ISD	-6.022 (-12.45, 0.41)	.067	.096	-4.644 (-10.994, 1.705)	.151	.094
Diastolic BP: Ref R2			.024			.024
Raw ISD	-0.003 (-0.010, 0.004)	.383	.024	-1.990 (-7.661, 3.681)	.491	.023
Res ISD	-2.233 (-6.275, 1.809)	.278	.025	-3.103 (-8.888, 2.683)	.293	.024
CoV	-4.286 (-13.73, 5.162)	.373	.024	-2.973 (-9.542, 3.596)	.374	.024
MSSD	-0.001 (-0.007, 0.004)	.676	.023	-1.279 (-6.869, 4.312)	.653	.023
Raw Fast	-0.009 (-0.049, 0.031)	.662	.023	2.417 (-2.884, 7.717)	.371	.024

ISD						
Res Fast ISD	-6.162 (-25.80, 13.47)	.538	.023	1.048 (-4.600, 6.696)	.716	.023
Raw Slow						
ISD	-0.006 (-0.013, 0.001)	.118	.027	-2.516 (-6.379, 1.347)	.201	.025
Res Slow						
ISD	-3.056 (-6.843, 0.730)	.113	.027	-2.257 (-5.993, 1.478)	.236	.025
Pulse: Ref R2			.006			.006
Raw ISD	0.007 (0.000, 0.014)	.047	.012	3.444 (-2.136, 9.024)	.226	.007
Res ISD	3.729 (-0.244, 7.702)	.066	.010	2.685 (-3.014, 8.385)	.355	.006
CoV	7.115 (-2.178, 16.41)	.133	.008	5.059 (-1.401, 11.520)	.125	.009
MSSD	**0.009 (0.003-0.014)**	**.001**	**.023**	6.837 (1.360, 12.313)	.015	.015
Raw Fast						
ISD	-0.019 (-0.059, 0.020)	.331	.006	-3.190 (-8.411, 2.024)	.230	.007
Res Fast ISD	-20.87 (-40.14, -1.60)	.034	.013	-6.699 (-12.23, -1.164)	.018	.015
Raw Slow						
ISD	0.005 (-0.002, 0.012)	.179	.017	1.515 (-2.294, 5.323)	.435	.005
Res Slow						
ISD	2.762 (-0.969, 6.492)	.146	.008	1.941 (-1.739, 5.621)	.301	.006
Peak Flow: Ref R2			.452			.452
Raw ISD	0.022 (-0.045, 0.090)	.520	.451	6.794 (-46.47, 60.06)	.802	.451
Res ISD	14.71 (-23.61, 53.03)	.451	.451	10.820 (-43.53, 65.17)	.696	.451
CoV	16.000 (-73.12, 105.11)	.724	.451	10.022 (-51.79, 71.84)	.750	.451
MSSD	0.032 (-0.020, 0.084)	.229	.452	9.822 (-42.80, 63.45)	.714	.451

Raw Fast						
ISD	0.078 (-0.302, 0.459)	.686	.451	26.394 (-24.25, 77.04)	.306	.452
Res Fast ISD	0.017 (-185.29,117.15)	.999	.451	7.000 (-46.56, 60.56)	.798	.451
Raw Slow						
ISD	0.009 (-0.058, 0.077)	.785	.451	4.324 (-31.83, 40.48)	.814	.451
Res Slow						
ISD	2.857 (-32.735, 38.448)	.875	.451	2.947 (-32.06, 37.95)	.869	.451
Grip Strength: Ref R2			.649			.649
Raw ISD	0.003 (-0.001, 0.008)	.173	.650	1.472 (-2.098, 5.042)	.418	.649
Res ISD	1.714 (-0.862, 4.290)	.192	.649	1.460 (-2.186, 5.107)	.432	.649
CoV	4.155 (-1.805, 10.116)	.171	.650	2.708 (-1.461, 6.877)	.203	.649
MSSD	0.002 (-0.001, 0.005)	.262	.649	1.039 (-2.475, 4.553)	.562	.649
Raw Fast						
ISD	-0.008 (-0.033, 0.017)	.513	.649	1.472 (-2.098, 5.042)	.418	.649
Res Fast ISD	-5.072 (-17.262, 7.118)	.414	.649	1.460 (-2.186, 5.107)	.432	.649
Raw Slow						
ISD	0.003 (-0.002, 0.007)	.235	.649	1.077 (-1.352, 3.507)	.384	.649
Res Slow						
ISD	1.497 (-0.883, 3.877)	.217	.649	1.260 (-1.098, 3.618)	.294	.649

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time; BP = blood pressure

RTI Scores in Relation to Change in Physical Functioning

SRT: The association between RTI scores for the SRT and longitudinal change in grip strength, peak flow, pulse, systolic blood pressure and diastolic blood pressure was examined using a series of mixed linear regression models with age, sex, education and mean reaction time on the SRT included as fixed effects, and age included as random effects. Results are presented in Table 6a. No significant associations were observed between SRT RTI scores and longitudinal change in physical functioning.

Lexical Decision: Mixed linear regression models with age, sex, education and mean reaction time on the Lexical Decision task included as fixed effects, and age included as a random effect, were used to examine the relationship between RTI scores for the Lexical decision test and longitudinal change in physical functioning. Results are presented in Table 6b. Significant associations were observed between RTI and longitudinal change in systolic blood pressure, such that higher log-transformed residual ISD scores were associated with less decline in systolic blood pressure. Untransformed raw ISD values computed from the slowest 20% of responses predicted longitudinal change in peak flow, but this association was also nonsignificant following log-transformation. No other significant associations were observed between Lexical RTI scores and longitudinal change in physical functioning.

Table 6a. Mixed linear regression of SRT RTI scores on objective measures of physical function.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	AIC	B (95% CI)	*p*	AIC
Systolic BP: Ref AIC			10283.13			10283.13
Raw ISD	0.003 (-0.002, 0.008)	.289	10299.66	0.896 (-0.459, 2.252)	.195	10276.72

Res ISD	0.348 (-0.230, 0.997)	.292	10280.31	0.887 (-0.469, 2.242)	.200	10268.75
CoV	1.079 (-0.659, 1.817)	.224	10275.86	0.917 (-0.454, 2.288)	.190	10276.64
MSSD	0.002 (-0.002, 0.006)	.271	10300.86	0.873 (-0.452, 2.198)	.196	10276.73
Raw Fast						
ISD	-0.007 (-0.04, 0.03)	.720	10292.66	-0.033 (-1.385, 1.319)	.962	10278.20
Res Fast						
ISD	-0.321 (-2.23, 1.59)	.742	10276.87	-0.029 (-1.391, 1.332)	.966	10278.14
Raw Slow						
ISD	0.001 (-0.002, 0.003)	.510	10302.89	0.223 (-0.528, 0.974)	.561	10280.20
Res Slow						
ISD	0.150 (-0.284, 0.584)	.497	10282.09	0.258 (-0.517, 1.032)	.514	10280.15
Diastolic BP: Ref AIC			8946.78			8946.78
Raw ISD	0.000 (-0.003, 0.003)	.759	8965.70	0.132 (-0.656, 0.920)	.743	8943.80
Res ISD	0.069 (-0.317, 0.436)	.756	8946.30	0.132 (-0.656, 0.920)	.742	8943.78
CoV	0.235 (-0.775, 1.245)	.648	8942.20	0.106 (-0.691, 0.902)	.795	8943.82
MSSD	0.001 (-0.001, 0.003)	.499	8967.13	0.294 (-0.476, 1.063)	.454	8943.68
Raw Fast						
ISD	0.007 (-0.014, 0.029)	.492	8957.92	0.429 (-0.356, 1.214)	.283	8943.05
Res Fast						
ISD	0.407 (-0.701, 1.516)	.471	8941.91	0.447 (-0.344, 1.237)	.268	8942.95
Raw Slow						
ISD	0.000 (-0.001, 0.002)	.503	8968.20	0.019 (-0.418, 0.455)	.933	8946.43
Res Slow	0.088 (-0.164, 0.340)	.494	8947.47	0.049 (-0.401, 0.499)	.832	8946.43

ISD						
Pulse: Ref AIC			9041.93			9041.94
Raw ISD	0.003 (-0.000, 0.006)	.079	9054.19	0.443 (-0.375, 1.261)	.289	9036.81
Res ISD	0.351 (-0.040, 0.741)	.078	9034.43	0.446 (-0.372, 1.264)	.285	9036.51
CoV	0.868 (-0.179, 1.914)	.104	9029.72	0.440 (-0.388, 1.267)	.298	9036.68
MSSD	0.003 (0.001, 0.005)	.016	9050.41	0.767 (-0.031, 1.564)	.060	9031.81
Raw Fast						
ISD	-0.012 (-0.03, 0.01)	.293	9052.52	-0.279 (-1.09, 0.54)	.503	9037.78
Res Fast						
ISD	-0.579 (-1.73, 0.57)	.324	9036.85	-0.262 (-1.08, 0.56)	.532	9037.69
Raw Slow						
ISD	0.001 (0.000, 0.003)	.048	9054.80	0.343 (-0.11, 0.79)	.138	9038.31
Res Slow						
ISD	0.262 (0.001, 0.523)	.049	9034.29	0.359 (-0.108, 0.827)	.131	9038.10
Peak Flow: Ref AIC			14221.31			14221.31
Raw ISD	0.002 (-0.026, 0.031)	.868	14231.98	-0.915 (-8.421, 6.591)	.811	14209.52
Res ISD	0.303 (-1.417, 2.022)	.730	16623.29	-0.963 (-8.47, 6.545)	.801	14209.55
CoV	-.111 (-9.804, 9.583)	.982	14208.73	-1.060 (-8.657, 6.537)	.734	14209.17
MSSD	0.004 (-0.017, 0.025)	.710	14233.15	0.000 (-7.348, 7.349)	.999	14209.35
Raw Fast						
ISD	-0.004 (-0.20, 0.19)	.968	14223.78	3.334 (-4.049, 10.71)	.376	14208.78
Res Fast						
ISD	-0.223 (-10.6, 10.16)	.967	14207.96	3.339 (-4.096, 10.77)	.378	14208.81

Raw Slow						
ISD	0.001 (-.013, 0.014)	.910	14234.35	-0.773 (-4.94, 3.393)	.716	14211.95
Res Slow						
ISD	0.157 (-2.302, 2.616)	.900	14213.47	-0.747 (-5.05, 3.551)	.733	14211.88
Grip Strength: Ref AIC			7678.41			7678.41
Raw ISD	-0.000 (-0.00, 0.00)	.899	7700.02	-0.052 (-0.56, 0.45)	.839	7677.58
Res ISD	-0.015 (-0.26, 0.23)	.905	7680.79	-0.052 (-0.56, 0.45)	.839	7677.60
CoV	-0.056 (-0.67, 0.59)	.865	7676.78	-0.072 (-0.58, 0.44)	.783	7677.54
MSSD	0.000 (-0.001, 0.001)	.920	7701.46	-0.022 (-0.52, 0.47)	.929	7677.71
Raw Fast						
ISD	-0.003 (-0.02, 0.011)	.672	7691.72	0.031 (-0.473, 0.536)	.903	7677.28
Res Fast						
ISD	-0.139 (-0.86, 0.58)	.703	7675.83	0.044 (-0.464, 0.552)	.864	7677.18
Raw Slow						
ISD	0.000 (-0.001, 0.001)	.735	7702.80	-0.013 (-0.29, 0.268)	.925	7679.61
Res Slow						
ISD	0.028 (-0.133, 0.189)	.732	7682.44	0.000 (-0.290, 0.290)	.999	7679.55

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time.

Table 6b. Mixed linear regression of Lexical Decision task RTI scores on change in physical function.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	AIC	B (95% CI)	*p*	AIC
Systolic BP: Ref AIC			10336.04			10336.05
Raw ISD	0.002 (-0.001, 0.004)	.197	10354.93	2.37 (0.498, 4.243)	.013	10322.40
Res ISD	0.913 (-0.432, 2.259)	.183	10329.18	**2.538 (1.637, 3.440)**	**.000**	**12783.14**
CoV	3.548 (0.506, 6.591)	.022	10322.45	2.458 (0.391, 4.525)	.020	10323.16
MSSD	0.000 (-0.001, 0.002)	.590	10357.59	1.762 (-0.032, 3.556)	.054	10323.89
Raw Fast ISD	-0.000 (-0.01, 0.01)	.982	10349.51	0.196 (-1.424, 1.816)	.813	10330.13
Res Fast ISD	0.011 (-6.167, 6.189)	.997	10324.39	0.151 (-1.548, 1.851)	.861	10330.10
Raw Slow ISD	0.002 (-0.001, 0.004)	.191	10351.79	1.319 (0.061, 2.578)	.040	10327.30
Res Slow ISD	**0.925 (0.279, 1.572)**	**.005**	**12786.78**	1.353 (0.147, 2.560)	.028	10326.92
Diastolic BP: Ref AIC			8986.78			8986.78
Raw ISD	0.001 (-0.001, 0.002)	.258	9008.05	0.976 (-0.110, 2.061)	.078	8979.10
Res ISD	0.459 (-0.321, 1.238)	.249	8982.24	1.024 (-0.086, 2.134)	.070	8979.41
CoV	1.580 (-0.183, 3.344)	.079	8977.82	1.158 (-0.041, 2.345)	.058	8978.66
MSSD	0.000 (-0.001, 0.001)	.684	9010.23	0.583 (-0.458, 1.624)	.272	8981.65
Raw Fast ISD	0.002 (-0.006, 0.009)	.648	9002.35	-0.015 (-0.456, 0.427)	.948	11441.34

Res Fast						
ISD	1.268 (-2.310, 4.846)	.487	8977.34	0.248 (-0.736, 1.231)	.621	8981.26
Raw Slow						
ISD	0.001 (-0.003, 0.003)	.148	9005.66	0.706 (-0.023, 1.435)	.058	8980.84
Res Slow						
ISD	0.581 (-0.214, 1.377)	.152	8980.50	0.784 (0.086, 1.483)	.028	8979.82
Pulse: Ref AIC			9088.91			9088.91
Raw ISD	-0.000 (-0.002, 0.00)	.602	9014.09	-0.454 (-1.588, 0.679)	.432	9084.04
Res ISD	-0.203 (-1.014, 0.61)	.623	9079.22	-0.421 (-1.590, 0.749)	.478	9084.08
CoV	-0.741 (-1.607, 0.13)	.094	11541.39	-0.476 (-1.726, 0.774)	.455	9082.91
MSSD	-0.001 (-0.002, 0.00)	.307	9092.49	-0.517 (-1.599, 0.566)	.349	9078.22
Raw Fast						
ISD	0.001 (-0.007, 0.008)	.890	9102.23	0.002 (-0.974, 0.979)	.996	9080.99
Res Fast						
ISD	0.911 (-2.809, 4.631)	.631	9073.37	0.287 (-0.735, 1.310)	.582	9077.14
Raw Slow						
ISD	0.000 (-0.002, 0.002)	.963	9106.36	-0.042 (-0.803, 0.718)	.913	9085.70
Res Slow						
ISD	0.135 (-0.693, 0.963)	.750	9079.15	0.128 (-0.601, 0.857)	.730	9083.70
Peak Flow: Ref AIC			14287.26			14287.26
Raw ISD	0.011 (-0.002, 0.024)	.087	14294.94	3.238 (-7.105, 13.58)	.539	14273.91
Res ISD	6.040 (-1.313, 13.39)	.107	14269.51	2.504 (-8.032, 13.04)	.641	14273.92
CoV	9.815 (-6.954, 26.58)	.251	14270.37	6.344 (-5.078, 17.76)	.276	14272.28

MSSD	0.008 (-0.001, 0.018)	.083	14292.67	5.681 (-4.200, 15.56)	.260	14264.85
Raw Fast						
ISD	0.007 (-0.062, 0.076)	.850	14286.19	-1.821 (-10.744, 7.10)	.689	14273.39
Res Fast						
ISD	6.058 (-27.82, 39.93)	.726	14269.43	0.207 (-9.163, 9.576)	.965	14274.70
Raw Slow						
ISD	**0.012 (0.006, 0.019)**	**.000**	**16716.47**	4.861 (-2.035, 11.75)	.167	14272.98
Res Slow						
ISD	6.145 (-1.36, 13.650)	.108	14271.79	3.350 (-3.268, 9.967)	.321	14274.95
Grip Strength: Ref AIC			7755.06			7755.06
Raw ISD	0.001 (0.000, 0.002)	.032	7759.08	0.571 (-0.138, 1.280)	.114	7742.43
Res ISD	0.553 (0.047, 1.060)	.032	7733.98	0.609 (-0.115, 1.332)	.099	7741.78
CoV	0.982 (-0.167, 2.13)	.094	7734.19	0.585 (-0.199, 1.369)	.143	7738.41
MSSD	0.001 (-0.000, 0.001)	.077	7764.27	0.540 (-0.140, 1.222)	.119	7744.16
Raw Fast						
ISD	0.001 (-0.003, 0.006)	.566	7771.44	0.037 (-0.576, 0.649)	.907	7753.39
Res Fast						
ISD	0.064 (-2.259, 2.386)	.957	7744.02	-0.174 (-0.817, 0.468)	.594	7750.71
Raw Slow						
ISD	0.001 (-0.000, 0.002)	.104	7767.31	0.233 (-0.245, 0.712)	.339	7749.09
Res Slow						
ISD	0.464 (-0.059, 0.980)	.082	7740.31	0.247 (-0.212, 0.706)	.291	7746.53

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time.

RTI Scores in Relation to Cognitive Functioning at Baseline

SRT: In a series of linear regression models adjusted for age, sex, education and mean reaction time, only two RTI scores for the SRT significantly predicted performance on tests in the VLS cognitive battery. A significant association was observed between performance on Similarities and log-transformed residual ISDs obtained from the fastest 20% of responses. The untransformed MSSD significantly predicted performance on the Letter Series test, but this association was not replicated for the log-transformed MSSD. Associations were negative such that increased RTI scores were associated with decreased cognitive performance. Raw and residual SRT ISDs yielded identical effect sizes in relation to all of the five cognitive test scores that were examined.

Table 7a. Linear regression of SRT RTI scores on baseline cognitive performance.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	R2 Adj	B (95% CI)	*P*	R2 Adj
Verbal Recall: Ref R2						
Raw ISD	-0.005 (-0.011, 0.002)	.142	.252	-1.306 (-3.086, 0.474)	.150	.252
Res ISD	-0.582 (-1.360, 0.196)	.142	.252	-1.336 (-3.118, 0.445)	.141	.252
CoV	-1.438 (-3.628, 0.752)	.198	.252	-1.390 (-3.183, 0.403)	.128	.252
MSSD	-0.004 (-0.008, 0.001)	.109	.253	-1.569 (-3.294, 0.155)	.074	.254

Raw Fast						
ISD	-0.041 (-0.088, 0.005)	.078	.253	-0.908 (-2.734, 0.917)	.329	.251
Res Fast						
ISD	-2.140 (-0.456-0.280)	.083	.253	-0.878 (-2.717, 0.961)	.349	.250
Raw Slow						
ISD	-0.001 (-0.004, 0.002)	.344	.257	-0.433 (-1.410, 0.544)	.384	.250
Res Slow						
ISD	-0.269 (-0.810, 0.273)	.330	.251	-0.462 (-1.464, 0.541)	.366	.250
Digit Symbol: Ref R2						
Raw ISD	-0.013 (-0.028, 0.001)	.070	.334	-4.923 (-9.071, 0.775)	.020	.336
Res ISD	-1.682 (-3.496, 0.132)	.069	.334	-4.946 (-9.098, 0.794)	.020	.336
CoV	-4.975 (-10.090, 0.141)	.057	.334	-5.041 (-9.220, 0.863)	.018	.336
MSSD	-0.008 (-0.018, 0.002)	.109	.333	-4.605 (-8.631, -0.58)	.025	.336
Raw Fast						
ISD	-0.114 (-0.221, -0.007)	.037	.335	-3.792 (-8.033, 0.448)	.080	.333
Res Fast						
ISD	-5.947 (-11.571, -0.324)	.038	.335	-3.823 (-8.092, 0.446)	.079	.333
Raw Slow						
ISD	-0.004 (-0.010, 0.003)	.305	.331	-1.325 (-3.606, 0.956)	.254	.331
Res Slow						
ISD	-0.642 (-1.906, 0.622)	.319	.331	-1.155 (-3.497, 1.188)	.333	.331
Similarities: Ref R2						
Raw ISD	0.001 (-0.008, 0.010)	.797	.117	-0.075 (-2.705, 2.556)	.955	.117

Res ISD	0.149 (-1.000, 1.299)	.799	.117	-0.098 (-2.731, 2.535)	.942	.117
CoV	0.473 (-2.761, 3.706)	.774	.117	0.106 (-2.544, 2.756)	.937	.117
MSSD	-0.001 (-0.008, 0.005)	.723	.117	-0.981 (-3.529, 1.567)	.450	.118
Raw Slow						
ISD	-0.069 (-0.137, -0.001)	.048	.123	-3.500 (-6.188, 0.812)	.011	.127
Res Slow						
ISD	-3.618 (-7.195, -0.040)	.048	.123	**-3.577 (-6.285, -0.87)**	**.010**	**.127**
Raw Fast						
ISD	0.001 (-0.003, 0.006)	.580	.117	0.408 (-1.037, 1.852)	.580	.117
Res Fast						
ISD	0.233 (-0.569, 1.035)	.568	.117	0.467 (-1.016, 1.949)	.537	.117
Letter Series: Ref R2						
Raw ISD	-0.007 (-0.013, -0.001)	.024	.291	-1.764 (-3.468, -0.06)	.042	.290
Res ISD	-0.890 (-1.665, -0.114)	.025	.291	-1.769 (-3.473, -0.07)	.042	.290
CoV	-0.236 (-4.489, -0.230)	.030	.291	-1.855 (-3.572, -0.14)	.034	.291
MSSD	**-0.006 (-0.010, -0.001)**	**.010**	**.293**	-1.889 (-3.537, -0.24)	.025	.291
Raw Fast						
ISD	-0.018 (-0.062, 0.026)	.426	.286	-0.929, (2.652, 0.794)	.290	.286
Res Fast						
ISD	-0.939 (-3.240, 1.362)	.423	.286	-0.919 (-2.655, 0.817)	.299	.286
Raw Slow						
ISD	-0.002 (-0.005, 0.001)	.121	.288	-0.511 (-1.446, 0.423)	.283	.286
Res Slow	-0.428 (-0.972, 0.116)	.123	.288	-0.537 (-1.497, 0.424)	.273	.286

ISD						
Vocabulary: Ref R2						
Raw ISD	-0.004 (-0.014, 0.005)	.381	.134	-0.510 (-3.287, 2.266)	.718	.133
Res ISD	-0.545 (-1.757, 0.668)	.378	.134	-0.537 (-3.316, 2.242)	.704	.133
CoV	-0.566 (-3.979, 2.848)	.745	.133	-0.602 (-3.399, 2.195)	.673	.133
MSSD	-0.003 (-0.010, 0.003)	.316	.135	-0.636 (-3.327, 2.054)	.643	.133
Raw Fast						
ISD	-0.033 (-0.105, 0.039)	.371	.134	-0.557 (-3.411, 2.297)	.702	.133
Res Fast						
ISD	-0.178 (-5.568, 2.007)	.356	.134	-0.607 (-3.482, 2.268)	.679	.133
Raw Slow						
ISD	0.000 (-0.005, 0.004)	.866	.133	-0.126 (-1.652, 1.399)	.871	.133
Res Slow						
ISD	-0.099 (-0.946, 0.747)	.818	.133	-0.225 (-1.790, 1.340)	.778	.133

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; SRT = simple reaction time.

Lexical Decision: In a series of linear regression models adjusted for age, sex, education and mean reaction time on the Lexical Decision Task, RTI scores for Lexical Decision were significantly associated with performance on Letter Series, Vocabulary, Similarities and Verbal Recall. Results are presented in Tables 7b. Significant associations were observed for raw and residual Lexical ISDs, the CoV, and the MSSD. All associations were negative, such that increased RTI scores were associated with decreased cognitive performance. Effect sizes were

similar across these RTI operationalizations, but the raw ISD tended to yield the strongest association with cognitive performance. Only one significant association was observed between lexical ISD scores calculated from the fastest 20% of responses and cognitive performance, such that the log-transformed residual ISD predicted performance on the Similarities test. Log transformation tended to increase the association between RTI scores and cognitive variables, with the largest increases in effect size occurring in relation to Verbal Recall and Similarities. Relative to the raw ISD, residual ISD and MSSD, log transformation of the CoV variable yielded less of an increase in effect size in relation to the observed cognitive variables.

Table 7b. Linear regression of Lexical RTI scores on baseline cognitive performance.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	R2 Adj	B (95% CI)	*P*	R2 Adj
Verbal Recall: Ref R2			.251			.251
Raw ISD	**-2.889 (-4.574, -1.204)**	**.001**	**.265**	**-5.454 (-7.779, -3.129)**	**.000**	**.278**
Res ISD	**-0.005 (-0.008, -0.002)**	**.001**	**.266**	**-5.406 (-7.783, -3.029)**	**.000**	**.276**
CoV	**-8.211 (-12.113, -4.31)**	**.000**	**.272**	**-5.729 (-8.430, -3.029)**	**.000**	**.273**
MSSD	**-0.003 (-0.005, -0.001)**	**.008**	**.260**	**-4.539 (-6.852, -2.226)**	**.000**	**.270**
Raw Fast ISD	0.020 (0.003, 0.037)	.019	.257	0.437 (-1.795, 2.670)	.700	.250
Res Fast ISD	8.779 (0.573-16.985)	.036	.256	-0.008 (-2.379, 2.363)	.995	.250
Raw Slow	**-0.006 (-0.009, -0.003)**	**.000**	**.268**	**-3.430 (-5.017, -1.844)**	**.000**	**.274**

ISD						
Res Slow						
ISD	**-2.924 (-4.500, -1.349)**	**.000**	**.268**	**-3.307 (-4.842, -1.771)**	**.000**	**.273**
Digit Symbol: Ref R2			.353			.353
Raw ISD	0.000 (-0.007, 0.007)	.967	.352	-1.658 (-7.090, 3.774)	.549	.353
Res ISD	0.533 (-3.363, 4.429)	.788	.352	-0.651 (-6.196, 4.894)	.818	.352
CoV	-0.812(-9.898-8.273	.861	.352	0.493 (-6.784, 5.797)	.878	.354
MSSD	-0.001 (-0.006, 0.004)	.629	.353	-3.040 (-8.408, 2.327)	.266	.354
Raw Fast						
ISD	-0.007 (-0.045, 0.031)	.717	.352	-2.909 (-8.011, 2.193)	.263	.354
Res Fast						
ISD	-0.221 (-19.08, 18.64)	.982	.352	-1.655 (-7.083, 3.774)	.550	.353
Raw Slow						
ISD	-0.003 (-0.010, 0.004)	.337	**.353**	-2.189 (-5.889, 1.511)	.246	.354
Res Slow						
ISD	-1.432 (-5.084, 2.220)	.441	.353	-1.487 (-5.063, 2.088)	.414	.353
Similarities: Ref R2			.142			.142
Raw ISD	**-0.007 (-0.011, -0.003)**	**.002**	**.156**	**-8.394 (-11.79, -5.00)**	**.000**	**.176**
Res ISD	**-3.721 (-6.190, -1.252)**	**.003**	**.154**	**-8.139 (-11.61, -4.67)**	**.000**	**.172**
CoV	**-12.123 (-17.83, -6.42)**	**.000**	**.167**	**-8.372 (-12.32, -4.42)**	**.000**	**.167**
MSSD	**-0.004 (-0.008, -0.001)**	**.009**	**.151**	**-7.142 (-10.52, -3.77)**	**.000**	**.167**
Raw Fast						
ISD	0.020 (-0.004, 0.044)	.109	.145	-2.031 (-5.29, 1.23)	.221	.143

Res Fast						
ISD	3.652 (-8.284, 15.589)	.548	.141	-3.748 (-7.190, -0.306)	.033	.148
Raw Slow						
ISD	**-0.009 (-0.014, -0.005)**	**.000**	**.167**	**-5.414 (-7.729, -3.100)**	**.000**	**.172**
Res Slow						
ISD	**-4.611 (-6.909, -2.312)**	**.000**	**.164**	**-4.814 (-7.060, -2.568)**	**.000**	**.167**
Letter Series: Ref R2			.313			.313
Raw ISD	**-0.006 (-0.008, -0.003)**	**.000**	**.331**	**-5.881 (-8.017, -3.745)**	**.000**	**.347**
Res ISD	**-2.911 (-4.466, -1.355)**	**.000**	**.329**	**-5.597 (-7.785, -3.409)**	**.000**	**.343**
CoV	**-8.688 (-12.286, -5.09)**	**.000**	**.339**	**-6.438 (-8.918, -3.959)**	**.000**	**.344**
MSSD	-0.003 (-0.005, -0.001)	.011	.320	**-4.656 (-6.787, 2.524)**	**.000**	**.335**
Raw Fast						
ISD	0.006 (-0.009, 0.022)	.428	.313	-0.961 (-3.026, 1.103)	.361	.313
Res Fast						
ISD	-1.662 (-9.316, 5.993)	.670	.312	**-3.092 (-5.279, -0.906)**	**.006**	**.322**
Raw Slow						
ISD	**-0.008 (-0.010, -0.005)**	**.000**	**.348**	**-4.331 (-5.777, -2.884)**	**.000**	**.353**
Res Slow						
ISD	**-3.923 (-5.356, -2.489)**	**.000**	**.347**	**-4.184 (-5.582, -2.786)**	**.000**	**.353**
Vocabulary: Ref R2			.210			.210
Raw ISD	**-0.013 (-0.018, -0.009)**	**.000**	**.256**	**-10.799 (-14.24, -7.36)**	**.000**	**.259**
Res ISD	**-7.200 (-9.682, -4.719)**	**.000**	**.252**	**-10.692 (-14.21, -7.17)**	**.000**	**.256**
CoV	**-17.73 (-23.49, -11.97)**	**.000**	**.257**	**-11.984 (-15.98, -7.99)**	**.000**	**.255**

MSSD	**-0.008 (-0.012, -0.005)**	**.000**	**.242**	**-9.922 (-13.34, -6.51)**	**.000**	**.252**
Raw Fast						
ISD	0.012 (-0.013, 0.037)	.341	.210	-1.187 (-4.531, 2.157)	.486	.209
Res Fast						
ISD	-0.014 (-0.019, -0.010)	.258	.210	-2.082 (-5.624, 1.460)	.249	.210
Raw Slow						
ISD	**-0.014 (-0.019, -0.010)**	**.000**	**.263**	**-6.833 (-9.185, -4.482)**	**.000**	**.252**
Res Slow						
ISD	**-7.352 (-9.663, -5.040)**	**.000**	**.261**	**-6.502 (-8.780, -4.225)**	**.000**	**.251**

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; BP = blood pressure.

RTI Scores in Relation to Change in Cognitive Functioning

SRT: Mixed linear regression models with age, sex, education and mean reaction time on the SRT included as fixed effects, and age included as a random effect were used to examine the relationship between RTI scores for the SRT and longitudinal change in cognitive functioning. Results are presented in Table 8a. No significant associations of baseline SRT RTI scores were observed with longitudinal change in any of the five cognitive measures. Associations remained null following log-transformation of SRT RTI scores.

Table 8a. Mixed linear regression of SRT RTI scores on longitudinal change in cognition.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	AIC	B (95% CI)	*p*	AIC

Verbal Recall: Ref AIC			7066.135			7066.135
Raw ISD	0.000 (-0.001, 0.001)	.900	7085.691	0.004 (-0.334, 0.342)	.980	7062.488
Res ISD	0.008 (-0.152, 0.170)	.910	7066.285	0.001 (-0.337, 0.339)	.995	7062.233
CoV	-0.009 (-0.441, 0.422)	.967	7062.197	0.015 (-0.327, 0.357)	.931	7062.283
MSSD	-0.000 (-0.001, 0.001)	.966	7085.874	0.005 (-0.324, 0.333)	.978	7059.723
Raw Fast						
ISD	0.001 (-0.008, 0.010)	.839	7079.633	-0.135 (-0.476, 0.206)	.437	7063.710
Res Fast						
ISD	0.043 (-0.438, 0.525)	.860	7063.887	-0.142 (-0.486, 0.201)	.417	7063.758
Raw Slow						
ISD	-0.000 (-0.001, 0.001)	.830	7089.590	-0.032 (-0.219, 0.155)	.736	7066.346
Res Slow						
ISD	-0.012 (-0.120, 0.095)	.822	7068.492	-0.042 (-0.234, 0.150)	.668	7065.32
Digit Symbol: Ref AIC			9080.453			9080.453
Raw ISD	-0.000 (-0.003, 0.003)	.850	9096.942	-0.089 (-0.878, 0.700)	.830	9068.731
Res ISD	-0.037 (-0.413, 0.340)	.849	9077.545	-0.086 (-0.875, 0.704)	.831	9068.739
CoV	-0.168 (-1.176, 0.841)	.744	9071.224	-0.067 (-0.865, 0.731)	.870	9068.834
MSSD	-0.000 (-0.002, 0.002)	.858	9099.456	-0.039 (-0.810, 0.732)	.921	9069.361
Raw Fast						
ISD	0.003 (-0.018, 0.025)	.289	9087.871	0.032 (-0.763, 0.827)	.937	9072.785
Res Fast						
ISD	0.168 (-0.956, 1.293)	.769	9072.234	0.048 (-0.752, 0.848)	.906	9073.113
Raw Slow	-0.000 (-0.002, 0.001)	.743	9103.234	-0.100 (-0.537, 0.340)	.659	9079.164

ISD						
Res Slow						
ISD	-0.042 (-0.295, 0.212)	.747	9082.507	-0.104 (-0.556, 0.348)	.651	9079.797
Similarities: Ref AIC			8100.763			8100.763
Raw ISD	0.001 (-0.000, 0.002)	.056	10618.173	0.201 (-0.309, 0.711)	.440	8098.794
Res ISD	0.243 (-0.861, 1.348)	.367	8100.329	0.202 (-0.309, 0.712)	.438	8098.804
CoV	0.231 (-0.419, 0.882)	.486	8097.114	0.182 (-0.334, 0.698)	.489	8098.658
MSSD	0.001 (-0.001, 0.002)	.458	8122.990	0.239 (-0.258, 0.735)	.346	8099.147
Raw Fast						
ISD	0.010 (-0.004, 0.023)	.170	8110.576	0.363 (-0.151, 0.877)	.166	8091.770
Res Fast						
ISD	0.518 (-0.208, 1.246)	.162	8094.659	0.374 (-0.143, 0.891)	.157	8091.410
Raw Slow						
ISD	0.000 (-0.001, 0.001)	.732	8123.243	-0.008 (-0.290, 0.274)	.958	8101.076
Res Slow						
ISD	0.027 (-0.136, 0.190)	.742	8102.414	-0.005 (-0.295, 0.286)	.976	8100.843
Letter Series: Ref AIC			6713.037			6713.037
Raw ISD	0.000 (-0.001, 0.002)	.547	6727.270	0.061 (-0.265, 0.388)	.710	6705.444
Res ISD	0.048 (-0.111, 0.208)	.550	6707.968	0.064 (-0.263, 0.391)	.702	6705.494
CoV	0.082 (-0.340, 0.505)	.703	6703.598	0.082 (-0.025, 0.412)	.628	6705.209
MSSD	0.000 (-0.001, 0.001)	.570	6726.379	0.080 (-0.239, 0.400)	.622	6704.470
Raw Fast						
ISD	0.001 (-0.008, 0.010)	.824	6727.078	-0.094 (-0.417, 0.230)	.570	6708.306

Res Fast						
ISD	0.053 (-0.405, 0.512)	.819	6711.225	-0.089 (-0.415, 0.237)	.592	6708.513
Raw Slow						
ISD	0.000 (-0.003, 0.001)	.518	6735.032	0.087 (-0.094, 0.268)	.346	6714.378
Res Slow						
ISD	0.036 (-0.072, 0.143)	.517	6714.215	0.090 (-0.096, 0.277)	.343	6714.169
Vocabulary: Ref AIC			8020.549			8020.549
Raw ISD	0.000 (-0.001, 0.002)	.625	8040.474	0.051 (-0.456, 0.558)	.840	8019.316
Res ISD	0.061 (-0.181, 0.304)	.622	8021.133	0.054 (-0.453, 0.561)	.835	8019.325
CoV	0.051 (-0.599, 0.701)	.878	8018.208	0.072 (-0.441, 0.584)	.784	8019.234
MSSD	-0.000 (-0.002, 0.001)	.870	8039.657	-0.105 (-0.599, 0.390)	.678	8017.574
Raw Fast						
ISD	0.010 (-0.003, 0.024)	.137	8031.999	0.090 (-0.415, 0.596)	.726	8019.537
Res Fast						
ISD	0.055 (-0.165, 1.259)	.132	8016.114	0.101 (-0.408, 0.610)	.697	8019.466
Raw Slow						
ISD	0.000 (-0.001, 0.001)	.839	8044.843	0.039 (-0.242, 0.320)	.786	8022.048
Res Slow						
ISD	0.020 (-0.143, 0.182)	.812	8023.909	0.051 (-0.239, 0.341)	.730	8021.831

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

Lexical Decision: Mixed linear regression models with age, sex, education and mean reaction time on the Lexical Decision Task included as fixed effects, and age included as a random effect were used to examine the relationship between RTI scores for the Lexical Decision task and longitudinal change in cognitive functioning. Results are presented in Table 8b. Of the eight RTI scores and five cognitive tests, only the Residual ISD obtained from the fastest 20% of responses significantly predicted increased decline in performance on the Similarities test. This association was no longer significant following log-transformation.

Table 8b. Mixed linear regression of Lexical RTI scores on longitudinal change in cognition.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*p*	AIC	B (95% CI)	*P*	AIC
Verbal Recall: Ref AIC			7094.431			7094.431
Raw ISD	0.000 (-0.000, 0.001)	.191	7104.860	0.384 (-0.080, 0.847)	.110	7069.532
Res ISD	0.211 (-0.124, 0.545)	.220	7080.444	0.384 (-0.090, 0.858)	.112	7071.125
CoV	0.612 (-0.136, 1.359)	.109	7070.916	0.460 (-0.051, 0.971)	.078	7072.565
MSSD	0.000 (-0.000, 0.001)	.236	7114.415	0.351 (-0.100, 0.802)	.127	7079.352
Raw Fast ISD	-0.002 (-0.006, 0.001)	.124	7108.364	-0.068 (-0.476, 0.341)	.746	7093.944
Res Fast ISD	-1.181 (-2.745, 0.381)	.138	7084.990	-0.131 (-0.561, 0.299)	.551	7093.618
Raw Slow ISD	0.000 (-0.000, 0.001)	.654	7102.458	0.234 (-0.076, 0.543)	.139	7075.147
Res Slow	0.093 (-0.238, 0.423)	.583	7078.263	0.272 (-0.026, 0.570)	.073	7076.905

ISD						
Digit Symbol: Ref AIC			9105.915			9105.915
Raw ISD	0.000 (-0.001, 0.002)	.706	9128.441	0.540 (-0.565, 1.644)	.340	9101.006
Res ISD	0.131 (-0.665, 0.927)	.747	9102.864	0.524 (-0.604, 1.651)	.362	9100.532
CoV	0.668 (-1.125, 2.460)	.465	9099.259	0.578 (-0.641, 1.797)	.352	9099.969
MSSD	0.000 (-0.001, 0.001)	.448	9129.352	0.600 (-0.462, 1.661)	.268	9100.741
Raw Fast						
ISD	0.001 (-0.006, 0.009)	.738	9121.091	0.189 (-0.763, 1.141)	.697	9099.311
Res Fast						
ISD	2.293 (-1.352, 5.939)	.217	9094.807	0.649 (-0.352, 1.651)	.204	9100.587
Raw Slow						
ISD	0.000 (-0.001, 0.002)	.526	9127.302	0.519 (-0.220, 1.258)	.168	9101.451
Res Slow						
ISD	0.124 (-0.685, 0.932)	.764	9102.455	0.296 (-0.414, 1.005)	.413	9102.720
Similarities: Ref AIC			8115.904			8115.904
Raw ISD	0.000 (-0.001, 0.001)	.546	8127.206	0.260 (-0.433, 0.953)	.460	8080.711
Res ISD	0.168 (-0.333, 0.670)	.511	8104.023	0.311 (-0.399, 1.021)	.390	8085.672
CoV	0.461 (-0.658, 1.581)	.419	8087.047	0.315810 (-0.45, 1.08)	.418	8086.506
MSSD	0.000 (-0.000, 0.001)	.533	8133.106	0.209 (-0.465, 0.883)	.543	8090.852
Raw Fast						
ISD	-0.002 (-0.007, 0.003)	.404	8131.658	0.145 (-0.467, 0.757)	.642	8113.585
Res Fast						
ISD	-1.466 (-2.57, -0.37)	.009	10615.626	0.051 (-0.592, 0.693)	.877	8107.738

Raw Slow						
ISD	0.000 (-0.001, 0.001)	.667	8120.102	0.141 (-0.322, 0.604)	.550	8085.970
Res Slow						
ISD	0.151 (-0.345, 0.647)	.551	8098.580	0.196 (-0.250, 0.642)	.389	8092.447
Letter Series: Ref AIC			6710.758			6710.758
Raw ISD	0.000 (-0.000, 0.001)	.284	6716.166	0.153 (-0.282, 0.588)	.490	6665.485
Res ISD	0.182 (-0.136, 0.499)	.261	6694.387	0.157 (-0.290, 0.603)	.491	6672.126
CoV	0.369 (-0.342, 1.080)	.308	6675.511	0.273 (-0.208, 0.754)	.266	6670.662
MSSD	0.000 (-0.000, 0.001)	.590	6729.797	0.018 (-0.402, 0.438)	.933	6678.687
Raw Fast						
ISD	-0.002 (-0.005, 0.001)	.110	6726.572	-0.130 (-0.513, 0.253)	.505	6704.567
Res Fast						
ISD	-0.663 (-2.131, 0.804)	.375	6701.505	0.071 (-0.331, 0.473)	.730	6696.647
Raw Slow						
ISD	0.000 (-0.000, 0.001)	.642	6696.308	0.087 (-0.202, 0.377)	.555	6660.217
Res Slow						
ISD	0.090 (-0.230, 0.409)	.582	6673.800	0.135 (-0.144, 0.414)	.343	6663.814
Vocabulary: Ref AIC			7935.393			7935.393
Raw ISD	0.000 (-0.001, 0.001)	.877	7878.402	0.421 (0.113, 0.728)	.018	10373.891
Res ISD	0.035 (-0.430, 0.500)	.881	7858.346	0.384 (-0.285, 1.05)	.260	7871.528
CoV	0.508 (-0.548, 1.563)	.345	7859.057	0.420 (-0.299, 1.140)	.252	7867.263
MSSD	0.000 (-0.001, 0.001)	.957	7899.760	0.356 (0.061, 0.651)	.018	10381.038
Raw Fast	-0.003 (-0.008, 0.001)	.161	7951.482	0.721 (-0.506, 0.650)	.807	7933.417

ISD						
Res Fast						
ISD	-1.727 (-3.920, 0.465)	.122	7926.197	-0.017 (-0.621, 0.587)	.956	7930.609
Raw Slow						
ISD	0.000 (-0.001, 0.001)	.950	7879.832	0.312 (-0.127, 0.751)	.164	7882.524
Res Slow						
ISD	0.015 (-0.459, 0.490)	.949	7857.770	0.302 (-0.120, 0.724)	.161	7884.118

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

RTI scores in Relation to Cognitive Status

SRT: Logistic regression models adjusted for age, sex, education and mean reaction time were used to examine the association between SRT RTI scores and four operationalizations of cognitive status. Results are presented in Table 9a. Only two significant associations were observed, such that untransformed raw and residual ISDs obtained from the fastest 20% of responses significantly predicted cognitive impairment operationalized by the presence of memory scores 1 SD below the mean. These associations were positive, such that increases in RTI predicted the presence of cognitive impairment. However, both associations were rendered nonsignificant following log-transformation.

Table 9a. SRT RTI scores as predictors of cognitive status.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*P*	ROC	B (95% CI)	*P*	ROC
MA-MCI, mean RT ROC:			0.599			.599
Raw ISD	0.997 (0.99-1.01)	0.997	0.566	0.805 (0.19-3.38)	.767	.566
Res ISD	0.649 (0.32-1.33)	0.239	0.564	0.789 (0.19-3.32)	.747	.564
CoV	0.350 (0.04-2.76)	0.318	0.526	0.758 (0.18-3.21)	.707	.526
MSSD	0.998 (0.99-1.01)	0.387	0.588	1.182 (0.31-4.57)	.809	.588
Raw Fast ISD	1.002 (0.969-1.035)	0.917	0.605	2.863 (0.695-11.804)	.146	.605
Res Fast ISD	1.137 (0.202-1.008)	0.884	0.607	2.918 (0.703-12.109)	.140	.607
Raw Slow ISD	0.998 (0.995-1.001)	0.296	0.512	0.779 (0.350-1.735)	.541	.512
Res Slow ISD	0.747 (0.431-1.292)	0.297	0.511	0.749 (0.328-1.711)	.493	.511
Mild single-domain MCI, mean RT ROC:			0.554			.554
Raw ISD	1.003 (0.998-1.008)	0.220	0.543	1.869 (0.399-8.755)	.427	.543
Res ISD	1.445 (0.805-2.592)	0.217	0.543	1.916 (0.409-8.986)	.410	.543
CoV	1.886 (0.299-11.902)	0.500	0.532	2.080 (0.443-0.976)	.354	.532
MSSD	1.002 (0.999-1.005)	0.186	0.562	2.414 (0.553-10.534)	.241	.562
Raw Fast ISD	**1.048 (1.014-1.084)**	**.006**	**.558**	1.925 (0.409-9.071)	.407	.558
Res Fast ISD	**11.282 (1.962-64.89)**	**.007**	**.558**	1.864 (0.392-8.864)	.434	.558
Raw Slow ISD	1.001 (0.999-1.003)	.337	.537	1.573 (0.666-3.715)	.301	.537
Res Slow ISD	1.222 (0.799-1.869)	.356	.528	1.514 (0.628-3.652)	.356	.528
Mild multi-domain MCI, mean RT ROC:			.613			.613
Raw ISD	1.004 (0.99-1.01)	.105	.602	3.46 (0.70-17.89)	.139	.602

Res ISD	1.639 (0.91-2.97)	.102	.602	3.626 (0.70-18.74)	.124	.602
CoV	3.278 (0.52-20.82)	.208	.569	3.64 (0.70-18.95)	.125	.569
MSSD	1.003 (1.00-1.01)	.093	.620	4.13 (0.86-19.86)	.077	.620
Raw Fast ISD	1.037 (1.001-1.074)	.043	.595	1.783 (0.325-9.784)	.505	.595
Res Fast ISD	6.295 (0.995-39.817)	.051	.594	1.696 (0.306-9.400)	.545	.594
Raw Slow ISD	1.002 (0.999-1.004)	.196	.575	2.116 (0.839-5.333)	.112	.575
Res Slow ISD	1.321 (0.859-2.032)	.204	.565	2.075 (0.810-5.310)	.129	.565
Moderate MCI, mean RT ROC:			.663			.663
Raw ISD	1.006 (1.00-1.01)	.039	.641	6.912 (0.950-50.313)	.056	.641
Res ISD	2.019 (1.038-3.926)	.038	.640	7.016 (0.963-51.136)	.055	.640
CoV	6.852 (0.870-53.986)	.068	.597	6.896 (0.936-50.810)	.058	.597
MSSD	1.004 (1.00-1.01)	.039	.656	7.622 (1.145-50.759)	.036	.656
Raw 20 ISD	1.030 (0.988-1.074)	.160	.659	2.842 (0.343-23.568)	.333	.659
Res 20 ISD	4.535 (0.505-40.691)	.177	.658	2.735 (0.327-22.867)	.353	.658
Raw 80 ISD	1.002 (1.000-1.005)	.060	.607	3.170 (1.018-9.869)	.046	.607
Res 80 ISD	1.564 (0.980-2.496)	.061	.599	3.220 (1.020-10.169)	.046	.599

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; ROC = receiver operating characteristic.

Lexical Decision: The same procedure was used to examine the association between Lexical Decision RTI scores and cognitive status. Results are presented in Table 9b. Of the untransformed RTI scores, only the CoV significantly predicted cognitive status as operationalized by MA-MCI. Neither the CoV nor any other untransformed RTI score predicted

any of the remaining three operationalizations of cognitive status. Following log-transformation, the raw ISD, residual ISD, CoV, MSSD and raw ISD scores associated with the fastest 20% of responses all significantly predicted MA-MCI, but associations with the other three operationalizations of cognitive status remained nonsignificant.

Table 9b. Lexical RTI scores as predictors of cognitive status.

	Raw RTI Scores			Log-Transformed RTI Scores		
	B (95% CI)	*P*	ROC	B (95% CI)	*P*	ROC
Mean RT, Ref ROC for MA-MCI:			.665			.665
Raw ISD	1.003 (1.000-1.005)	.016	.663	**18.731 (2.907-120.692)**	**.002**	**.663**
Res ISD	3.961 (1.229-12.764)	.021	.662	**18.173 (2.789-118.409)**	**.002**	**.662**
CoV	**61.151 (3.740-999.73)**	**.004**	**.635**	**19.675 (2.243-172.608)**	**.007**	**.635**
MSSD	1.001 (1.000-1.003)	.053	.652	**11.998 (2.025-71.084)**	**.006**	**.652**
Raw Fast ISD	1.002 (0.990-1.014)	.759	.669	**13.240 (2.149-81.561)**	**.005**	**.669**
Res Fast ISD	0.075 (0.000-25.656)	.385	.622	3.379 (0.538-21.225)	.194	.622
Raw Slow ISD	1.002 (1.00-1.004)	.026	.635	3.752 (1.052-13.380)	.042	.635
Res Slow ISD	3.645 (1.243-10.686)	.018	.642	4.400 (1.259-15.378)	.020	.642
Mild single-domain MCI, mean RT ROC:			.618			.618
Raw ISD	1.002 (1.000-1.004)	.035	.619	6.103 (0.878-42.439)	.068	.619

Res ISD	3.810 (1.143-12.706)	.029	.622	7.507 (1.052-53.551)	.044	.622
CoV	23.580 (1.086-512.02)	.044	.607	8.399 (0.801-88.132)	.076	.607
MSSD	1.001 (1.000-1.003)	.105	.621	5.766 (0.873-38.070)	.069	.621
Raw Fast						
ISD	1.001 (1.000-1.003)	.345	.590	1.311 (0.179-9.632)	.790	.590
Res Fast						
ISD	0.011 (0.000-9.076)	.190	.574	0.916 (0.118-7.099)	.916	.574
Raw Slow						
ISD	1.001 (0.999-1.003)	.243	.588	1.620 (0.414-6.333)	.488	.588
Res Slow						
ISD	2.020 (0.703-5.806)	.192	.602	2.181 (0.575-8.274)	.252	.602
Mild multi-domain MCI, mean RT ROC:						.670
Raw ISD	1.002 (1.000-1.005)	.042	.666	13.161 (0.154-112.204)	.018	.666
Res ISD	3.890 (1.097-13.789)	.035	.670	16.064 (1.841-140.42)	.012	.670
CoV	44.74 (0.161-124.434)	.025	.641	16.873 (1.227-231.992)	.035	.641
MSSD	1.001 (0.999-1.003)	.212	.660	8.174 (1.043-64.028)	.045	.660
Raw Fast						
ISD	0.994 (0.979-1.008)	.381	.641	2.831 (0.303-26.483)	.362	.641
Res Fast						
ISD	0.004 (0.000-5.320)	.132	.603	1.027 (0.104-10.129)	.981	.603
Raw Slow						
ISD	1.001 (0.999-1.003)	.204	.631	2.641 (0.586-11.913)	.206	.631
Res Slow	2.142 (0.711-6.453)	.176	.645	3.240 (0.740-14.183)	.119	.645

ISD						
Moderate MCI, mean RT ROC:						.662
Raw ISD	1.003 (1.001-1.006)	.016	.663	17.967 (1.304-247.627)	.031	.663
Res ISD	6.719 (1.560-28.946)	.011	.673	26.048 (1.777-381.748)	.017	.673
CoV	130.100 (2.54-660.1)	.015	.639	28.520 (1.158-72.696)	.040	.639
MSSD	1.002 (1.000-1.004)	.098	.653	9.510 (0.808-111.887)	.073	.653
Raw Fast						
ISD	0.991 (0.974-1.007)	.271	.607	1.403 (0.092-21.430)	.808	.607
Res Fast						
ISD	0.010 (0.000-46.158)	.287	.584	0.853 (0.051-14.147)	.912	.584
Raw Slow						
ISD	1.002 (1.000-1.004)	.086	.636	3.895 (0.633-23.976)	.143	.636
Res Slow						
ISD	3.025 (0.899-10.179)	.074	.646	1.403 (0.092-21.430)	.808	.646

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency; ROC = receiver operating characteristic.

Results Summary

We observed skewness and kurtosis in the distribution of SRT and Lexical Decision RTI scores in this sample, which prompted a replication of analyses using log-transformed RTI scores. Log-transformation significantly modified associations between RTI scores and physical and cognitive outcomes, especially for those scores with large initial deviations from normality.

For example, as illustrated in Table 2, ISDs computed from the fastest and slowest 20% of responses tended to deviate furthest from normality, and these scores also yielded associations with study outcomes that were least resilient to the effects of log-transformation. The results summarized here reflect findings associated with log-transformed RTI scores.

Overall, relative to the SRT, RTI scores obtained from the Lexical Decision task shared stronger and more consistent associations with measures of both physical and cognitive functioning. Associations between RTI and measures of physical function were weak and inconsistent overall. Of the measures examined, only cross-sectional pulse and longitudinal change in systolic blood pressure and peak flow were associated with Lexical RTI. The observed associations between RTI and change in physical function should not be overstated because none of these findings were observed across untransformed and log-transformed computations of RTI. In addition, mixed regression models from which these results were obtained did not include baseline RTI scores as a random effect. With respect to cognitive function, RTI scores obtained from the SRT again shared weak and inconsistent associations with the tests in the VLS battery. In contrast, RTI on the Lexical Decision task was significantly associated with baseline performance on tests of memory and reasoning. No RTI scores were significantly associated with processing speed performance. Neither the SRT nor the Lexical Decision task RTI scores yielded consistent or meaningful associations with cognitive change.

When examined in relation to cognitive status, RTI scores obtained from the SRT did not reliably predict cognitive impairment. In contrast, RTI scores obtained from the Lexical Decision task yielded a consistent association with cognitive impairment as operationalized by MA-MCI. Of the eight Lexical RTI scores that were examined, the raw ISD, the CoV, the residual ISD and the MSSD yielded comparable associations with cognitive status.

Study 1 Discussion

Many approaches to operationalizing RTI have been proposed and implemented in the field of cognitive aging. However, to date there has been no direct comparison of the associations among various RTI computations or of their relative associations with physical and cognitive outcomes of interest. This study examined eight operationalizations of RTI obtained from two different RT tasks: the SRT, a measure of simple reaction time, and the Lexical Decision task, a measure of choice reaction time. The main finding of this study regarding the relative sensitivity of these eight RTI scores was that across both the SRT and the Lexical Decision task, the raw ISD, residual ISD, CoV and MSSD yielded comparable associations with physical and cognitive outcomes. There was a tendency for the Residual ISD to yield slightly stronger associations with study outcomes, but the relative advantage of the Residual ISD relative to the other operationalizations was minimal (e.g., Rsquare differences < 0.03). Several authors have advocated for the use of the residual ISD on conceptual grounds (e.g., Hultsch, MacDonald & Dixon, 2002), and substantial evidence in older adults and other populations suggests that the residual ISD is sensitive to diseases of aging. (e.g., Dykiert, Der, Starr, & Deary, 2012). This study provides further support for the utility of the residual ISD as a measure of RTI for use in aging research, but did not observe an advantage of the residual ISD relative to the raw ISD, CoV or MSSD.

The CoV has been widely used in the aging literature as an index of variability. Research examining the CoV in relation to imaging outcomes has found it to perform equivalent to the Ex-Gaussian tau parameter (Jackson, Balota, Duchek, & Head, 2012), and it has been found in a prospective study of risk factors for mortality to yield highly similar associations with mortality as the raw and residual ISD (Batterham, Bunce, Mackinnon & Christensen, 2014). The results of

the current study provide further support for a strong association between CoV and ISD-based operationalizations of RTI. Associations with cognitive outcomes tended to be most significant using ISD-based operationalizations of RTI, but there was no evidence of any systematic discrepancy of findings between ISDs and CoV that might implicate differences at the construct level. Thus the current study concludes that the CoV is a valid and feasible approach to operationalizing variability that adjusts for the relation between RTI and mean RT.

In the current literature, the unadjusted ISD, the mean-adjusted ISD and the CoV are most prominently represented computations of RTI. The results of this study provide support for the utility of all three of these operationalizations of RTI. The remaining operationalizations that were examined in this study, the MSSD and ISD scores obtained from "fast" and "slow" responses, were included on more conceptual grounds. "Fast" and "slow" ISDs have been examined previously in the literature, and associations with cognitive performance in these studies have predominantly been documented in ISDs obtained from the slowest tail of intraindividual responses (Hultsch, MacDonald & Dixon, 2002). Results of this study support the hypothesis that associations between RTI and cognitive function are based on RTI in the slow tail of the distribution of responses. However, ISDs calculated from the slow tail did not yield stronger associations with cognitive outcomes than were observed for ISDs obtained from the full distribution. Further, ISDs calculated from the slow tail of responses were found to be more vulnerable to the effects of log-transformation than the ISDs calculated from the full distribution. On the basis of these findings, ISDs computed from the full distribution were concluded to be more reliable operationalizations of RTI that yield associations of an equal magnitude with physical and cognitive outcomes.

The MSSD has been used extensively for quantification of variability in blood pressure, cardiac function, and bold response. However, its application in behavioural research is more limited. By operationalizing variability based on sequential changes in reaction time, the MSSD naturally adjusts for trial-based changes in reaction time due to learning effects or fatigue. Although, unlike the residual ISD and CoV, it does not adjust for mean RT, results of the current study suggest that the MSSD is equivalent to ISDs and the CoV in its association with cognition. Not all operationalizations of RTI are represented in this work. In particular, distribution-based operationalizations such as are obtained from Ratcliff and ex-Gaussian parameters could not be obtained because these require more trials per participant than were available from VLS data.

The current study examined RTI scores calculated from the SRT and the Lexical Decision Task. Prior research has found that cognitively demanding RT tasks yield stronger associations with cognitive outcomes (e.g., Gorus, De, Lambert, Lemper, & Mets, 2008). However, robust associations have also been reported between RTI obtained from the SRT and CNS integrity in older adults (e.g., Cherbuin, Sachdev, & Anstey, 2010). The current study supported the finding that RTI scores obtained from cognitively demanding tasks yield stronger associations with cognitive performance. RTI scores computed from the SRT did not yield reliable associations with any of the study outcomes. In contrast, RTI scores computed from the Lexical Decision task significantly predicted the MA-MCI classification of cognitive status, and predicted performance on four of the five cognitive tests in the VLS battery.

The present study found no consistent association between RTI and measures of physical functioning. Prior research examining associations between physical function and RTI has been mixed. Anstey and colleagues (2005) examined associations between RTI and performance on objective measures of physical functioning and found a significant association between forced

expiratory volume and RTI such that higher RTI predicted lower expiratory volume. It is noteworthy that in the current data, unadjusted associations between RTI in the SRT and Lexical Decision task, but these were rendered nonsignificant after adjusting for age and mean reaction time. Thus age effects and individual differences in mean RT appear to account for associations between physical functioning and RTI in this generally healthy, high functioning sample.

Several studies have documented significant associations between RTI and longitudinal change in cognitive function such that increased RTI predicts greater cognitive decline. In particular, MacDonald and colleagues (2003) found significant coupling between increases in RTI and decreases in cognitive performance over time in data from the VLS (Macdonald, Hultsch, & Dixon, 2003). Similar findings have been reported in other datasets (e.g., Bielak, Strauss, MacDonald, Hultsch & Hunter, 2010). Time-lagged analyses by Lovden and colleagues (2007) extended these findings by demonstrating that changes in RTI in cognitively stable individuals are predictive of subsequent cognitive change (Lovden, Li, Shing, & Lindenberger, 2007). However, both of these investigations primarily observe correlated change in RTI and cognition, rather than a significant effect of baseline RTI for predicting longitudinal cognitive change. Lovden and colleagues, who observed correlated change in RTI and cognition, in fact found that baseline RTI did not independently predict cognitive change in their sample. In addition, the findings reported by MacDonald and colleagues were obtained from composite scores obtained from RTI scores computed from four RT tasks of varying complexity. In particular, RTI in performance on episodic and semantic recognition tasks contributed to their RTI composite. It is possible that the discrepancy between results of the current study and those reported by MacDonald and colleagues (2003) is a reflection of the added cognitive demands of tasks that contributed to their composite score.

Reaction time has been shown to slow with increasing age (e.g., Der & Deary, 2006). It has thus been suggested that associations between RTI and physical and cognitive function in older adults may be an artefact of mean RT rather than a property of RT with unique predictive value (Salthouse, 2012). Faster examinees are indeed much more likely to produce smaller RTI values and slower examinees are more likely to produce larger RTI values. In the current study, correlations between mean RT and RTI confirmed the presence of moderate to high associations between these scores, even for the residual ISD from which age, sex and trial effects were residualized prior to analysis. As properties of the same distribution of RT values, it may not be surprising that an association exists between mean RT and RTI. However, findings reported in this study regarding the association between RTI and physical and cognitive outcomes were all obtained from models that included mean RT, participant age, and sex as covariates. Thus, these results all represent unique associations between RTI and physical and cognitive outcomes independent of mean RT. Concern has been expressed in the literature regarding the potential for systematic individual differences in the relation between mean RT and RTI to bias results of RTI analyses that adjust for mean RT though covariation or use of the CoV (Schmiedek, Lovden & Lindenberger, 2009; Hultsch, Strauss, Hunter & MacDonald, 2007). There is no consensus in the literature regarding the optimal approach to disentangling effects of mean RT from those associated with RTI, but studies directly comparing basic distribution-based operationalizations of RTI (e.g., CoV, ISD) with more computationally intensive approaches that better isolate mean and RTI (e.g., ex-Gaussian parameters) have observed highly similar associations with outcomes relevant to CNS function (Batterham, Bunce, Mackinnon & Christensen, 2014; Jackson, Balota, Duchek, & Head, 2012).

There was substantial variability in the strength of the mean-RTI association across the eight examined computations of RTI. The most marked delineation in relative association with mean RT was observed between the SRT ISDs computed from the fastest vs. slowest 20% of responses. Mean RT shared a strong association with variability in the fast tail of the RT distribution, but only a weak association with variability in the slow tail of the RT distribution. The reverse pattern was observed in bivariate associations between the whole-distribution operationalizations of RTI (raw ISD, residual ISD, CoV, MSST) and ISDs computed from the fast and slow tails of the RT distribution. Whole-distribution RTI scores correlated very highly with RTI in the slow tail of responses, but only shared small to moderate associations with RTI in the fast tail of responses. In addition, there was virtually no association between RTI in the fast vs. slow tails of the RT distribution. Though less pronounced than the SRT, the described pattern of associations was also observed for Lexical Decision task. Taken together these results suggest that inconsistency in the fast and slow tails of the RT distribution may reflect distinct aspects of RTI. This is further supported by the different patterns of associations with physical and cognitive functioning that were observed for the fast vs. slow ISD scores. Prior research has demonstrated that the sensitivity of RTI to age and neurological integrity is a reflection of increased slow responses (Hultsch, Macdonald, & Dixon, 2002). The correlations among mean RT and each of the examined RTI scores support this notion, but further suggest that increases in task complexity may attenuate correlations between RTI scores and slow RT values.

The current study took between-person and within-person steps to ensure that RT outliers were excluded from each examinee's distribution of RT values. This was done because extreme responses, usually falling in the slow tail of the RT distribution, can inflate RTI scores and may introduce external sources of measurement error. First, consistent with prior research, absolute

values were established for “extreme” fast and slow responses for the SRT and the Lexical Decision task. All scores for all examinees that fell beyond these limits were deleted. Second, at the within-person level, all RT values falling more than three standard deviations beyond each examinee’s intraindividual mean were deleted. Identification of outliers using the intraindividual mean RT is preferable to identification of outliers using the group mean because the group mean RT can result in over-detection of extreme responses in slower examinees and under-detection in fast examinees (Salthouse, 2012).

In summary, the present study identified four operationalizations of RTI that, when obtained from Lexical Decision RT data, are essentially interchangeable in their positive associations with cognitive status. These scores will be used in subsequent analyses to determine whether their documented association with cognitive status in the participants of the VLS is retained following standardization against an independent normative sample.

Study 2: Norm-Referenced Operationalizations of RTI: Predictive Validity for Detecting Cognitive Impairment in Older Adults

With population aging and corresponding projected increases in the prevalence of cognitive impairment and dementia in the adult population, there is keen interest in the identification of brief, valid methods for early detection of cognitive impairment. A premise of the emphasis on early detection is that it will aid identification of candidates for clinical trials and allow at-risk individuals maximal time to address modifiable risk factors (e.g., vascular disease). As is reviewed in this dissertation, there is a large body of literature suggesting that RTI provides a sensitive measure of cognitive dysfunction in older adults. When examined in relation to Alzheimer's disease, RTI has been shown to increase in correspondence with disease severity (Gorus, De, Lambert, Lemper, & Mets, 2008), and to reliably predict the presence of AD biomarkers in healthy examinees (Duchek et al., 2009). Associations between RTI and cognitive status have also been demonstrated using as few as 40 trials of an RT task (Anstey et al., 2007), which satisfies the practical demand for brief and easily administered tests. Thus, preliminary evidence supports investigation of the potential clinical utility of measures of RTI.

One obstacle to the clinical extension of RTI relates to the current lack of consensus regarding best practices for operationalizing RTI. Many different methods for computing RTI are reported in the literature, and there is limited evidence that these RTI computations are interchangeable or even that they measure the same central construct. Study 1 addressed this issue by examining associations between eight operationalizations of RTI on RT tasks of increasing complexity in relation to cognitive status and measures of cognitive and physical functioning in participants of the Victoria Longitudinal Study (VLS). Consistent with prior research, results of Study 1 suggested that the Lexical Decision task yielded RTI scores that

shared stronger associations with cognitive status than RTI scores obtained from the SRT. Of the eight operationalizations of RTI examined, the raw ISD, the residual ISD, the CoV and the MSSD were all found to yield similar associations with cognitive performance and cognitive status. Thus, these computations appear to be comparable in their sensitivity to cognitive functioning in older adults. The finding of a stronger association between with Lexical ISD and cognitive outcomes relative to the SRT is consistent with prior research demonstrating that the association between RTI and cognitive outcomes increased as a function of the complexity of the RT task from which RTI scores were computed.

The clinical assessment of psychological processes and cognitive abilities relies on normative comparison. Norm-referenced testing involves comparing the observed test score of an examinee to the scores of a sample of individuals who are comparable on critical dimensions to the examinee. Norm-referenced testing relies on the assumption that the abilities measured by a given test are normally distributed in the populations (Nunnally & Bernstein, 1994). It is expected that most people, 68.26% of all examinees, will perform within one standard deviation of the mean of their normative stratum. Individuals who score more than one standard deviation above or below the mean are classified as having higher or lower aptitude for the abilities measured by a given test relative to others their age. Sufficiently extreme responses, usually classified as 1.5 to 2 standard deviations below the mean, are classified as impaired and considered to reflect a pathologically low aptitude for the abilities measured by a given test relative to other examinees with similar demographic characteristics. To date there has been no published attempt to determine whether norm-referencing of RTI scores can be used to detect cognitive impairment in older adults, or whether the same distribution-based classifications of impairment can be applied to RTI scores. In order to address this question, the current study

replicated positive results from Study 1 after standardizing RTI scores against demographically stratified RTI data from an independent sample of VLS participants.

The finding reported in Study 1 of positive linear associations between RTI and cognitive status does not itself demonstrate the clinical utility of RTI. For example, the finding that individuals with the high RTI scores experience poorer cognitive function than individuals with low RTI scores is important from a proof-of-concept perspective, but to compare one exceptional group to another is not clinically meaningful. In order to have potential clinical utility, it is necessary for individuals with RTI scores in the average range to serve as the reference groups for statistical comparison. Thus, in addition to replicating findings from Study 1, Study 2 examined whether VLS participants with norm-referenced RTI scores in the high range experienced poorer cognitive function and a higher prevalence of cognitive impairment relative to participants with scores in the average range. Study 2 further examined whether individuals with lower levels of RTI experienced higher cognitive function and a lower prevalence of impairment relative to those in the average range.

Another issue that must be addressed in the evaluation of the clinical utility of a test relates to classification accuracy. The mere presence of a statistically significant association between a measure and an outcome provides little insight into clinical utility (Akobeng, 2007). Likelihood ratios (LR), calculated by obtaining the ratio of a test's true positives relative to false positives, are among the most popular statistics to inform clinical decisions. LRs provide information about the diagnostic utility of a test based on the modification the test result would make to the pretest probability of the presence of the outcome in a given examinee (Akobeng, 2007). Established guidelines are available that can then be followed to determine whether a test's contribution to prediction of an outcome is meaningfully different from the pretest

probability. Thus, another objective of Study 2 was to examine results in relation to established guidelines for clinically meaningful LR values. In summary, the purpose of the current study was 1) to replicate findings from Study 1 using norm-referenced RTI scores, 2) to examine whether distribution-based assumptions of standardized assessment can be applied to RTI scores, and 3) to examine whether associations between RTI and cognitive status are of a sufficient magnitude to meaningfully contribute to clinical decision making.

Study 2 Methods

The Victoria Longitudinal Study (VLS) is a longitudinal study of multiple facets of human aging. Details of the VLS are presented in greater detail in the Methods section of Study 1 (pg. 15) and elsewhere (Dixon & de Frias, 2004; Hultsch, Hertzog, Dixon & Small, 1998). The current study recruited data from Sample 2 of the VLS to assemble normative data for those RTI operationalizations that were found in Study 1 to predict cognitive status. Sample 3, Wave 1 RTI data were then standardized against Sample 2 normative data and the obtained T-Scores were examined in relation to the physical and cognitive outcomes that are described in Study 1. The measures of physical and cognitive function that were examined in this study are described in Study 1 (pg. 15-17, pg.18-20). RTI scores were computed as described in Study 1 (pg. 17-18).

Development of Normative Data

The sequential cohort design of the VLS was capitalized upon in the creation of normative RTI data by standardizing Sample 3, Wave 1 RTI Scores against Sample 2, Wave 1 data. Tasks, trials and instructions associated with the reaction time measures were identical across samples and occasions. RTI scores from all Sample 2 Wave 1 examinees except for those meeting criteria for MA-MCI were compiled to serve as normative data for Sample 3. Linear regression was then used to determine the extent to which demographic stratification of the Sample 2 RTI data was necessary. Age, sex and educational attainment were examined together in a linear regression model along with all associated interaction terms. In order to determine the optimal level of age stratification for Sample 2 RTI normative data, categorical variables corresponding to various levels of age stratification (e.g., 2-year vs. 5-year vs. 10-year stratification) were examined in relation to RTI scores to determine which level of stratification, expressed in terms of pseudo R-square, retained the strongest association with RTI. Gender

(male vs. female) and education (12 or fewer years of education vs. more than 12 years of education) were examined as dichotomous variables. Education was recorded as a continuous variable in the VLS, but was examined as a dichotomous variable for the purpose of stratifying the normative sample. Further stratification was not possible due to sample size limitations.

Operationalization of RTI

RTI scores that yielded the strongest and most reliable associations with physical and cognitive outcomes in Study 1 were examined in the present study. Thus, the raw ISD, residual ISD, CoV and MSSD values obtained from the Lexical Decision task were examined in this study. The computation of these scores followed the methods described in Study 1 (pg. 17-18). As deviations from normality were observed in Study 1 to compromise the stability of study findings, all normative data from Sample 2 and RTI data from Sample 3 were log-transformed prior to examination. Log-transformation compromises the feasibility of RTI for clinical use by adding to the computational intensiveness of calculating RTI scores, but was deemed necessary because distribution-based classifications of normal functioning vs. impairment assume the presence of a normal distribution of data (Nunnally & Bernstein, 1994).

Statistical Analyses

Linear Regression Analyses: Linear regression was first used in the process of assembling Sample 2 normative data to examine the relationship between RTI scores and relevant demographic variables. Following standardization of Sample 3 RTI data, obtained T-Scores were examined in relation to continuous physical and cognitive outcomes. The univariate associations between RTI T-Scores and cognitive and physical outcomes were examined first in linear regression models with a given RTI T Score as the sole predictor. Subsequently, RTI T-Scores were examined in relation to physical and cognitive performance in models adjusting for

mean RT, age, sex and education. Demographic covariates were included in the adjusted models because, although the RTI T-Scores were adjusted for relevant demographic characteristics, the mean RT and outcome variables are unadjusted. To account for multiple comparisons, only *p*-values < 0.01 were interpreted as statistically significant.

Binomial Logistic Regression Analyses: Binomial logistic regression analyses were used to examine the association between RTI T-Scores and cognitive status in the VLS sample. The classification of cognitive status in the present study followed the methods described in Study 1 (pg. 18-19). As with the linear regression analyses, the predictive value of RTI T-Scores for detecting cognitive impairment was examined first in a series of univariate logistic regression models, and subsequently in models adjusted for Mean RT, age, sex and education. To account for multiple comparisons, only *p*-values < 0.01 were interpreted as statistically significant.

Multinomial Logistic Regression Analyses: Multinomial logistic regression analyses were conducted to establish whether distribution-based assumptions of standardized assessment could be applied to the study of RTI. Specifically, these analyses addressed whether, relative to individuals with RTI T-Scores falling in the average range, individuals with elevated RTI T-Scores exhibited poorer physical and cognitive performance and a higher prevalence of cognitive impairment. Similarly, these analyses examined whether RTI T-Scores in the low range were associated with higher physical and cognitive performance and a lower prevalence of impairment relative to individuals with RTI T-Scores in the average range. This is an important extension of the binomial logistic regression analyses described above because it is more clinically meaningful to compare the likelihood of cognitive impairment in individuals with high vs. average RTI (as is accomplished in the multinomial analysis) than to compare the likelihood of cognitive impairment in individuals with low vs. high RTI (as is assessed by analyses where RTI

scores are included as continuous predictors). For the current analyses, T-Score values of 65 or higher were classified as "high" RTI, and values below 45 were classified as "low" RTI. The classification of values of 65 and higher as "high" RTI is in keeping with conventional assessment cutoffs that classify all values falling 1.5 SD beyond the mean as clinically elevated. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Stratum-Specific Likelihood Ratios: Stratum-specific likelihood ratios (SSLRs) are an extension of conventional likelihood ratios that determine the clinical implications of performing within a given range on a diagnostic test. SSLRs are obtained by calculating the ratio of the proportion of total cases performing within a given range to the proportion of total healthy controls performing in the same range. Established guidelines for interpreting likelihood ratios can be used to interpret the degree to which performing within a given range on a diagnostic test makes a clinically meaningful contribution to the detection of a disease. Likelihood ratios between 1-2 and between 0.5-1 are said to alter probabilities by a "rarely important" range, those falling between 2-5, and between 0.2-0.5, alter probabilities by small but "sometimes important" magnitudes, those between 5-10 and between 0.1-0.2 alter probabilities by a "moderate" magnitude, and those greater than 10 or less than 0.1 alter probabilities by "large and often conclusive" magnitudes (Jaeschke, Guyatt & Sackett, 1994).

Study 2 Results

Assembly and demographic stratification of VLS normative data

RT data from VLS Sample 2 participants who were cognitively intact at baseline were used to compute normative RTI data for use in subsequent analyses. Demographic characteristics of the normative sample relative to the experimental sample are presented in Table 10.Groups differed significantly only in their performance on the Vocabulary test, in which the experimental group performed a mean of two points lower than the normative group. The raw ISD, residual ISD, CoV and MSSD were computed in the normative sample following the same procedures applied to the experimental sample (described in Study 1, pg. 17-18). All RTI scores were log-transformed to account for the deviations from normality reported in Study 1 (pg. 24-25). To determine the degree of stratification that was necessary in order to obtain unbiased standardized RTI scores, log-transformed RTI data were examined for associations with age, sex and education using linear regression. Results are presented in Table 11. Across all four examined operationalizations of RTI in the Lexical Decision task, age and education were significant predictors of RTI and sex yielded no independent association with RTI. These findings prompted the stratification of VLS normative data by age and educational attainment. Optimal levels of age stratification were tested empirically by examining 5 and 10-year categorical age variables in relation to RTI scores using linear regression. Although both variables yielded significant associations with RTI, the 5-year categorical variable accounted for more variance in RTI than the 10-year variable (Continuous age variable R-square =0.08; 5-year age variable R-square = 0.06; 10-year age variable R-square = 0.05). Thus, sample 2 RTI normative data were stratified in 5-year intervals. Due to sample size limitations, educational

stratification involved obtaining RTI estimates separately from examinees who reported 12 years or less, and those reporting more than 12 years of education.

Table 10. Demographic Characteristics of VLS Normative and Experimental Samples.

	Normative	Experimental	F (df), *p
N	462	577	
Age	68.48 (7.56)	68.30 (8.60)	0.13 (1), *NS*
Sex	88%	68%	0.17 (1), *NS*
Education	14.83 (3.06)	15.17 (2.99)	3.25 (1), *NS*
Verbal Recall	17.73 (4.34)	17.31 (4.49)	2.31 (1), *NS*
Letter Series	11.12 (4.55)	11.57 (4.29)	2.68 (1), *NS*
Similarities	16.18 (6.22)	15.73 (6.17)	1.35 (1), *NS*
Vocabulary	**44.06 (7.15)**	**42.33 (6.60)**	**16.35 (1)****
Digit Symbol	49.28 (11.41)	49.39 (11.02)	0.03 (1), *NS*
Peak Flow	417.08 (116.06)	418.55 (116.91)	0.04 (1), *NS*
Grip Strength	30.91 (9.70)	31.05 (9.71)	0.05 (1), *NS*
Mean Lexical	1124.32 (432.83)	1089.42 (403.56)	1.80 (1), *NS*

Note. VLS = Victoria Longitudinal Study; df = degrees of frequency; NS = nonsignificant. All values are presented as mean (standard deviation) unless noted otherwise.

Table 11. Linear associations between demographic variables and RTI Scores

	Raw ISD	Residual ISD	CoV	MSSD
Age	**8.89 (5.97, 11.82)****	**0.02 (0.01, 0.02)****	**0.002 (0.001, 0.003)****	**12.60 (8.53, 16.66)****
Education	**-8.62 (-15.96, -1.28)***	**-0.01, (-0.03, -0.00)***	**-0.003 (-0.006, -0.001)***	**-11.91 (-22.09, -1.72)***
Sex	-5.01 (-52.39, 42.37) *NS*	-0.01 (-0.09, 0.07) *NS*	-0.010 (-0.028, 0.008) *NS*	-12.33 (-78.13, 53.47) *NS*

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; * denotes *p*-values < 0.05; ** denotes *p*-values < 0.01.

Univariate characteristics of standardized RTI scores

RT data from Sample 3 participants of the VLS were standardized against the Sample 2 normative data and converted to T-Scores. Distributional characteristics of the Sample 3 standardized RTI T Scores are presented in Table 12a. All T-Scores were normally distributed with mean values approximating 50 and standard deviations approximating 10. Bivariate associations among RTI T-Scores were comparable to the previously reported bivariate associations among raw Sample 3 RTI values (e.g., pg. 26).

Table 12a. Univariate characteristics of RTI T-Scores in the VLS Sample

Lexical T-Score	Mean	SD	Skewness	SE Skew	Kurtosis	SE Kurtosis
Raw ISD	50.843	11.143	0.510	0.103	0.495	0.206
Residual ISD	50.968	10.975	0.686	0.103	0.811	0.206
CoV	50.905	11.072	0.327	0.103	0.669	0.206
MSSD	48.345	10.690	0.577	0.103	0.498	0.206

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; SD = standard deviation; SE = standard error.

Bivariate Associations among RTI Scores

Bivariate associations among mean RT and the four operationalizations of RTI in Sample 2 (normative sample) and Sample 3 are presented in Table 12b. Associations with mean RT and correlations among the four operationalizations of RTI were virtually identical in Sample 2 and Sample 3 VLS data.

Table 12b. Correlation matrices for Lexical RTI Scores in VLS Samples 2 and 3

	Sample 2 Lexical RTI Scores				Sample 3 Lexical RTI Scores			
	Raw ISD	Res ISD	CoV	MSSD	Raw ISD	Res ISD	CoV	MSSD
Lexical								
Mean	**.81**	**.81**	**.50**	**.80**	**.81**	**.82**	**.51**	**.81**
Raw ISD	**1.00**	**.99**	**.89**	**.99**	**1.00**	**.99**	**.90**	**.96**
Res ISD		**1.00**	**.89**	**.99**		**1.00**	**.89**	**.96**
CoV			**1.00**	**.88**			**1.00**	**.83**
MSSD				**1.00**				**1.00**

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

As described in the Methods, a categorical score was computed from RTI T-Scores by classifying scores as falling within “Low”, “Average” and “High” ranges. Characteristics of participants falling within each of these T-Score Strata are presented in Table 13. There were no significant differences in age, sex or objective physical functioning of participants across low, average and high RTI strata for any of the four RTI computations. The prevalence of cognitive impairment (as operationalized by MA-MCI) was significantly higher in participants with RTI scores in the “High” range. There was also a strong consistent association between mean reaction time on the Lexical Decision task and RTI stratum across all four computations of RTI.

Table 13. Characteristics of Sample 3 VLS participants by RTI Strata

	Low (162)	Average (311)	High (48)	*F* (*df*), *p*
Raw ISD Strata				
Age	67.93 (7.89)	68.38 (8.94)	69.00 (8.58)	0.342 (2) *NS*
Sex, % F	70	66	71	0.564 (2) *NS*
Education	15.51 (2.77)	15.18 (2.94)	14.43 (3.41)	2.63 (2) *NS*
Cognition, % impaired	**5**	**14**	**29**	**10.72 (2)****
Systolic BP	126.06 (17.72)	126.02 (15.98)	127.30 (14.48)	0.329 (2) *NS*
Grip Strength	31.20 (9.16)	31.10 (9.80)	31.03 (10.45)	0.036 (2) *NS*
Mean Lexical	**820.14 (137.83)**	**1119.79 (264.80)**	**1842.61 (700.4)**	**225.06 (2)****
Residual ISD Strata				
Age	67.79 (8.03)	68.22 (9.0)	69.19 (8.51)	1.07 (2) *NS*
Sex, % F	68	68	69	0.025 (2) *NS*
Education	**15.54 (2.79)**	**15.19 (2.98)**	**14.19 (3.25)**	**5.24 (2)****
Cognition, % impaired	**6**	**13**	**27**	**9.33 (2)****
Systolic BP	127.24 (17.97)	125.40 (15.86)	127.66 (14.13)	1.26 (2) *NS*
Grip Strength	31.20 (9.16)	31.10 (9.80)	31.03 (10.45)	0.083 (2) *NS*
Mean Lexical	**813.14 (136.45)**	**1113.15 (253.501)**	**1827.94 (700.06)**	**174.26 (2)****
CoV Strata				
Age	69.49 (9.22)	67.69 (8.52)	66.93 (7.37)	3.04 (2)*
Sex, % F	64	70	68	0.938 (2) *NS*

Education	15.42 (2.88)	15.25 (2.85)	14.37 (3.65)	2.26 (2) *NS*
Cognition, %				
impaired	**8**	**12**	**29**	**8.78 (2)****
Systolic BP	128.04 (18.32)	125.23 (15.81)	126.20 (13.33)	1.577 (2) *NS*
Grip Strength	31.09 (9.11)	31.86 (9.73)	32.66 (10.62)	0.937 (2) *NS*
Mean Lexical	**908.87 (252.37)**	**1110.21 (400.87)**	**1440.33 (480.3)**	**45.59 (2)****
MSSD Strata				
Age	68.60 (8.71)	67.92 (8.87)	67.20 (6.78)	0.933 (2) *NS*
Sex, % F	67	68	71	0.182 (2) *NS*
Education	15.62 (2.80)	15.01 (2.94)	14.39 (3.71)	2.716 (2) *NS*
Cognition, %				
impaired	**5**	**15**	**29**	**13.87 (2)****
Systolic BP	125.66 (16.55)	126.44 (16.59)	126.02 (14.50)	0.045 (2) *NS*
Grip Strength	31.40 (9.22)	30.82 (9.76)	31.93 (11.23)	0.335 (2) *NS*
Mean Lexical	**854.66 (165.67)**	**1169.72 (304.28)**	**1804.03 (775.2)**	**173.47 (2)****

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; * denotes *p*-values < 0.05; ** denotes *p*-values < 0.01.

RTI T-Scores in relation to cognitive performance

Linear regression models were used to examine the relationship between RTI T-Scores and performance on the five cognitive tests in the VLS battery. These associations were examined both using univariate regression models with RTI T-Scores as the sole predictors of

cognitive performance, and subsequently using models adjusted for age, sex, education and mean RT. Results are presented in Table 14. Significant negative associations were observed between all four RTI scores and four of the five cognitive tests in the VLS battery, such that higher RTI T-Scores predicted lower cognitive performance. Only the CoV did not significantly predict performance on the Digit Symbol test. Univariate associations of RTI with cognition were small, accounting for 3-15% of variability in performance on the five cognitive tasks. Associations were weakest for Digit Symbol and strongest for Vocabulary. The raw ISD and residual ISD tended to yield stronger associations with cognitive performance than the CoV or MSSD. Adjustment for mean RT and demographic covariates rendered nonsignificant associations between RTI and performance on the Digit Symbol test, but did not affect the significance or direction of associations between RTI T-Scores and other cognitive tests.

Given the significant independent effects of RTI T-Scores in the linear analyses, multinomial regression was used to examine whether cognitive performance differed in individuals with “high” or “low” RTI T-Scores relative to those with scores in the average range. The categorical variable classifying RTI T-Scores as “low”, “average” or “high” was used as the outcome of univariate multinomial logistic models with cognitive test scores included as predictors. Results are presented in Table 15. Relative to participants with RTI scores in the average range, cognitive performance was lower in individuals with high RTI scores, and higher in individuals with low RTI. Exceptions were observed for the Digit Symbol test, which did not differ significantly across levels of CoV scores, and did not differ significantly in examinees with residual ISD scores in the low range relative to the average range. Following mean adjustment, only Vocabulary and the Digit Symbol retained significant associations with low or high RTI scores relative to RTI scores in the average range.

Table 14. Linear regression of Lexical RTI T Scores on VLS cognitive tests.

	Unadjusted			Adjusted		
	B (95% CI)	*P*	R2 Adj	B (95% CI)	*P*	R2 Adj
Linear regression of Lexical ISD T Score on Verbal Recall Test						
Raw ISD	**-0.113 (-0.145, -0.081)**	**.000**	**.077**	**-0.099 (-0.143, -0.055)**	**.000**	**.262**
Res ISD	**-0.097 (-0.130, -0.064)**	**.000**	**.055**	**-0.089 (-0.133, -0.044)**	**.000**	**.257**
CoV	**-0.088 (-0.120, -0.055)**	**.000**	**.045**	**-0.071 (-0.104, -0.037)**	**.000**	**.259**
MSSD	**-0.096 (-0.130, -0.062)**	**.000**	**.051**	**-0.083 (-0.128, -0.037)**	**.000**	**.253**
Linear regression of Lexical ISD T Score on Digit Symbol (Correct)						
Raw ISD	**-0.247 (-0.327, -0.167)**	**.000**	**.061**	-0.009 (-0.111, 0.094)	.868	.348
Res ISD	**-0.188 (-0.270, -0.105)**	**.000**	**.030**	0.008 (-0.094, 0.111)	.871	.348
CoV	-0.102 (-0.185, -0.019)	.016	.009	0.006 (-0.071, 0.083)	.873	.348
MSSD	**-0.213 (-0.298, -0.129)**	**.000**	**.041**	-0.040 (-0.145, 0.065)	.456	.349
Linear regression of Lexical ISD T Score on Similarities						
Raw ISD	**-0.153 (-0.196, -0.109)**	**.000**	**.080**	**-0.164 (-0.227, -0.101)**	**.000**	**.187**
Res ISD	**-0.140 (-0.185, -0.095)**	**.000**	**.061**	**-0.141 (-0.205, -0.078)**	**.000**	**.177**
CoV	**-0.127 (-0.172, -0.082)**	**.000**	**.051**	**-0.097 (-0.144, -0.049)**	**.000**	**.173**
MSSD	**-0.143 (-0.189, -0.096)**	**.000**	**.060**	**-0.140 (-0.205, -0.075)**	**.000**	**.176**
Linear regression of Lexical ISD T Score on Letter Series						
Raw ISD	**-0.125 (-0.155, -0.094)**	**.000**	**.103**	**-0.100 (-0.141, -0.059)**	**.000**	**.335**
Res ISD	**-0.108 (-0.140, -0.077)**	**.000**	**.075**	**-0.096 (-0.137, -0.055)**	**.000**	**.332**
CoV	**-0.101 (-0.133, -0.070)**	**.000**	**.066**	**-0.105 (-0.131, -0.078)**	**.000**	**.326**

MSSD	**-0.107 (-0.140, -0.075)**	**.000**	**.070**	**-0.108 (-0.136, -0.080)**	**.000**	**.326**
Linear regression of Lexical ISD T Score on Vocabulary						
Raw ISD	**-0.228 (-0.273, -0.183)**	**0.000**	**0.150**	**-0.210 (-0.275, -0.145)**	**.000**	**.243**
Res ISD	**-0.236 (-0.281, -0.190)**	**0.000**	**0.154**	**-0.194 (-0.259, -0.128)**	**.000**	**.235**
CoV	**-0.220 (-0.265, -0.174)**	**0.000**	**0.137**	**-0.151 (-0.200, -0.102)**	**.000**	**.239**
MSSD	**-0.240 (-0.286, -0.193)**	**0.000**	**0.152**	**-0.192 (-0.260, -0.125)**	**.000**	**.232**

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; * denotes *p*-values < 0.05; ** denotes *p*-values < 0.01.

Table 15. Multinomial regression analysis of cognitive performance by RTI strata.

	Unadjusted		Mean-Adjusted	
	Low Avg.	High	Low Avg.	High
Raw ISD Lexical Decision Strata				
Recall	**1.109 (1.059, 1.161)****	**0.898 (0.841, 0.958)****	1.007 (0.948, 1.069) *NS*	0.989 (0.907, 1.078) *NS*
Letters	**1.094 (1.043, 1.147)****	**0.852 (0.794, 0.915)****	0.947 (0.886, 1.012) *NS*	0.986 (0.901, 1.079) *NS*
Similarities	**1.069 (1.036, 1.103)****	**0.931 (0.882, 0.982)****	1.004 (0.963, 1.046) *NS*	0.996 (0.931, 1.046) *NS*
Vocabulary	**1.098 (1.058, 1.139)****	**0.909 (0.874, 0.945)****	1.052 (1.004, 1.102)*	0.946 (0.897, 0.997)*
Digit Symbol	**1.025 (1.007, 1.043)****	**0.942 (0.915, 0.970)****	**0.961 (0.986, 0.990)****	1.013 (0.997, 1.050) *NS*
Residual ISD Lexical Decision Strata				
Recall	**1.087 (1.040, 1.137)****	**0.909 (0.853, 0.970)****	0.975 (0.920, 1.034) *NS*	0.992 (0.917, 1.072) *NS*
Letters	**1.085 (1.036, 1.137)****	**0.901 (0.841, 0.964)****	0.932 (0.874, 0.995)*	1.032 (0.949, 1.121) *NS*
Similarities	**1.046 (1.014, 1.078)****	**0.916 (0.869, 0.966)****	0.971 (0.932, 1.012) *NS*	0.968 (0.909, 1.030) *NS*
Vocabulary	**1.082 (1.043, 1.121)****	**0.897 (0.862, 0.932)****	1.031 (0.986, 1.078) *NS*	**0.924 (0.880, 0.970)****
Digit Symbol	1.015 (0.998, 1.032) *NS*	**0.949 (0.923, 0.976)****	**0.941 (0.916, 0.967)****	1.008 (0.976, 1.041) *NS*
Lexical Decision CoV Strata				

Recall	**1.070 (1.024, 1.118)****	0.924 (0.869, 0.982)*	1.021 (0.973, 1.072) *NS*	0.963 (0.901, 1.029) *NS*
Letters	**1.092 (1.041, 1.145)****	**0.911 (0.853, 0.972)****	1.034 (0.980, 1.090) *NS*	0.964 (0.898, 1.035) *NS*
Similarities	1.035 (1.004, 1.067)*	**0.925 (0.880, 0.972)****	1.003 (0.970, 1.037) *NS*	0.947 (0.899, 0.998)*
Vocabulary	**1.083 (1.045, 1.123)****	**0.920 (0.886, 0.955)****	**1.056 (1.017, 1.097)****	**0.938 (0.901, 0.977)****
Digit Symbol	1.003 (0.987, 1.021) *NS*	0.987 (0.962, 1.012) *NS*	**0.970 (0.950, 0.991)****	1.021 (0.993, 1.051) *NS*
Lexical Decision MSSD Strata				
Recall	**1.096 (1.051, 1.142)****	0.947 (0.881, 1.019) *NS*	0.989 (0.037, 1.044) *NS*	1.052 (0.962, 1.151) *NS*
Letters	**1.104 (1.057, 1.154)****	0.919 (0.851, 0.992)*	0.971 (0.915, 1.030) *NS*	1.055 (0.961, 1.157) *NS*
Similarities	**1.072 (1.041, 1.105)****	0.941 (0.886, 0.999)*	1.015 (0.977, 1.053) *NS*	0.992 (0.927, 1.063) *NS*
Vocabulary	**1.114 (1.076, 1.153)****	**0.923 (0.884, 0.963)****	**1.079 (1.035, 1.126)****	0.959 (0.911, 1.010) *NS*
Digit Symbol	**1.024 (1.008, 1.041)****	**0.958 (0.928, 0.988)****	**0.961 (0.938, 0.985)****	1.021 (0.985, 1.059) *NS*

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; * denotes p-values < 0.05; ** denotes p-values < 0.01.

RTI T-Scores in relation to cognitive status

Binomial logistic regression models were used to examine the sensitivity of Lexical RTI T-Scores to the four examined operationalizations of cognitive status. These associations were examined first using univariate models with RTI Scores as the sole predictor of cognitive status, and subsequently using adjusted models with age, sex, education and mean RT included as covariates. Results are presented in Table 16. Consistent positive associations were observed in the unadjusted models such that higher RTI T-Scores significantly increased the probability of cognitive impairment. In the adjusted models these associations attenuated for all operationalizations of cognitive status except the MA-MCI. The residual ISD and the MSSD T-Scores tended to yield slightly stronger associations with cognitive status relative to the raw ISD and the CoV. Across all four operationalizations of cognitive status, ROC curve values for the residual ISD and the MSSD were stronger than the ROC value for mean RT. The residual ISD had 71% sensitivity and 55% specificity when the cutoff value of the predicted probability of impairment was 0.11.

Table 16. Logistic analyses of RTI T-Scores as predictors of cognitive status.

	Unadjusted			Mean-Adjusted	
	B (95% CI)	*P*	ROC	B (95% CI)	*P*
Multi-Assessment MCI					
Ref (Mean RT)			0.665	**1.001 (1.000, 1.001)**	**0.002**
Raw ISD	**1.051 (1.029, 1.075)**	**0.000**	**0.655**	**1.054 (1.020, 1.090)**	**0.002**
Residual ISD	**1.054 (1.032, 1.078)**	**0.000**	**0.672**	**1.057 (1.025, 1.091)**	**0.000**
CoV	**1.052 (1.028, 1.077)**	**0.000**	**0.641**	**1.043 (1.017, 1.070)**	**0.001**

MSSD	**1.056 (1.032, 1.080)**	**0.000**	**0.669**	**1.058 (1.024, 1.093)**	**0.001**
Memory impairment -1.0 SD and no other test score below -0.50.					
Ref (Mean RT)			0.618	**1.001 (1.000, 1.001)**	**0.004**
Raw ISD	**1.037 (1.013, 1.062)**	**0.003**	**0.613**	1.023 (0.988, 1.059)	0.204
Residual ISD	**1.040 (1.016, 1.064)**	**0.001**	**0.626**	1.028 (0.995, 1.063)	0.098
CoV	**1.037 (1.012, 1.063)**	**0.003**	**0.614**	1.026 (0.999, 1.054)	0.068
MSSD	**1.041 (1.016, 1.067)**	**0.001**	**0.632**	1.029 (0.995, 1.065)	0.100
Memory impairment -1.5 SD or greater plus a score of -1.0 SD on at least one other test.					
Ref (Mean RT)			0.670	**1.001 (1.000, 1.002)**	**0.001**
Raw ISD	**1.051 (1.025, 1.079)**	**0.000**	**0.663**	1.037 (0.998, 1.078)	0.061
Residual ISD	**1.052 (1.026, 1.079)**	**0.000**	**0.676**	1.039 (1.002, 1.077)	0.038
CoV	**1.048 (1.020, 1.076)**	**0.001**	**0.648**	1.033 (1.003, 1.064)	0.033
MSSD	**1.052 (1.025, 1.081)**	**0.000**	**0.672**	1.037 (0.998, 1.077)	0.060
Memory impairment -2.0 SD or greater plus a score of -1.5 SD on at least one other test.					
Ref (RT Mean)			0.662	**1.001 (1.000, 1.002)**	**0.008**
Raw ISD	**1.055 (1.022, 1.088)**	**0.001**	**0.657**	1.050 (1.002, 1.099)	0.039
Residual ISD	**1.061 (1.029, 1.094)**	**0.000**	**0.692**	**1.062 (1.016, 1.109)**	**0.007**
CoV	**1.058 (1.024, 1.093)**	**0.001**	**0.663**	1.047 (1.010, 1.085)	0.013
MSSD	**1.060 (1.026, 1.094)**	**0.000**	**0.678**	1.056 (1.009, 1.105)	0.019

Note. *Note.* RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; ROC = receiver operator characteristic.

Multinomial regression was then used to examine whether the prevalence of cognitive impairment differed significantly in individuals with "high" or "low" RTI T-Scores relative to those with RTI scores in the average range. Results are presented in Table 17. All four operationalizations of RTI yielded significant associations with MA-MCI such that individuals with RTI scores in the high range experienced higher odds of cognitive impairment than those in the Average range, and those with RTI scores in the low range experienced lower odds of impairment. Associations with the remaining operationalizations were more variable. Stratum-specific likelihood ratios were calculated to determine the clinical significance of the associations between RTI T-Scores and MA-MCI. As is presented in Table 18, SSLRs indicated that the presence of an RTI T-Score in the low range, regardless of the RTI operationalization used, can make a small but potentially clinically meaningful contribution to the process of ruling out the presence of cognitive impairment. Similarly, scores in the high range, regardless of RTI operationalization, are associated with a small but potentially clinically meaningful increase in the odds of cognitive impairment.

Table 17. Prevalence of cognitive impairment by RTI T-Score stratum.

	Low	Avg.	High
Raw ISD Lexical Decision Strata			
MA-MCI	**0.345 (0.165, 0.723)****		**2.530 (1.265, 5.062)****
Memory -1 SD	0.680 (0.342, 1.353) *NS*		2.658 (1.242, 5.690)*
Memory -1.5 SD	0.378 (0.158, 0.929)*		2.714 (1.229, 5.994)*
Memory - 2 SD	0.310 (0.090, 1.067) *NS*		2.944 (1.161, 7.466)*
Residual ISD Lexical Decision Strata			

MA-MCI	**0.394 (0.194, 0.803)****	**2.447 (1.227, 4.881)****
Memory -1 SD	0.580 (0.287, 1.173) *NS*	2.150 (0.989, 4.673) *NS*
Memory -1.5 SD	0.291 (0.111, 0.763)*	2.160 (0.960, 4.858) *NS*
Memory - 2 SD	0.205 (0.047, 0.896)*	**3.365 (1.378, 8.218)****
CoV Lexical Decision Strata		
MA-MCI	**0.208 (0.093, 0.468)****	**0.338 (0.173, 0.658)****
Memory -1 SD	**0.289 (0.113, 0.738)****	0.554 (0.257, 1.192) *NS*
Memory -1.5 SD	**0.183 (0.057, 0.585)****	0.634 (0.276, 1.457) *NS*
Memory - 2 SD	**0.108 (0.028, 0.425)****	0.321 (0.131, 0.784)*
MSSD Lexical Decision Strata		
MA-MCI	**0.300 (0.154, 0.581)****	**2.523 (1.187, 5.364)****
Memory -1 SD	**0.353 (0.180, 0.690)****	0.975 (0.359, 2.649) *NS*
Memory -1.5 SD	**0.194 (0.080, 0.470)****	0.857 (0.287, 2.561) *NS*
Memory - 2 SD	**0.169 (0.050, 0.572)****	1.507 (0.488, 4.652) *NS*

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference.

Table 18. Stratum-specific likelihood ratios associated with RTI T-Score strata.

Raw ISD Cutoff	MCI	No MCI	%MCI	% no MCI	SSLR
Low	9	165	0.1304348	0.3360489	**0.388142**
Average	46	291	0.6666667	0.592668	1.124857
High	14	35	0.2028986	0.0712831	**2.846377**
Res ISD Cutoff	MCI	No MCI	%MCI	% no MCI	SSLR
Low	10	164	0.1449275	0.3333333	0.434783
Average	45	291	0.6521739	0.5914634	1.102645
High	14	37	0.2028986	0.0752033	**2.698002**
CoV Cutoff	MCI	No MCI	%MCI	% no MCI	SSLR
Low	10	156	0.1492537	0.3170732	0.470723
Average	45	296	0.6716418	0.601626	1.116378
High	12	40	0.1791045	0.0813008	**2.202985**
MSSD Cutoff	MCI	No MCI	%MCI	% no MCI	SSLR
Low	12	219	0.173913	0.4460285	**0.389915**
Average	45	246	0.6521739	0.5010183	1.301697
High	12	26	0.173913	0.0529532	**3.284281**

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; MCI = mild cognitive impairment; SSLR = stratum-specific likelihood ratio.

Study 2 Discussion

Many investigators have reported elevated RTI in individuals with behavioural or imaging markers of CNS dysfunction (e.g., Jackson, Balota, Duchek, & Head, 2012; Gorus, De, Lambert, Lemper, & Mets, 2008). Given the importance of identifying behavioural methods for early detection of cognitive impairment, the current study sought to determine whether standardized measures of RTI would conform to the assumptions of standardized assessment, and whether the strength of associations between standardized measures of RTI and cognitive status was sufficient for clinical utility. The main finding of this study was that RTI T-Scores bear many of the qualities of a measure that is appropriate for clinical use.

RTI T-Scores were found to share associations with cognitive performance and cognitive status that were of a similar direction and magnitude as were observed of raw RTI scores in Study 1. Associations were strongest for the VLS measures of memory, vocabulary and reasoning, and tended to be weaker for digit symbol, a measure of processing speed and complex attention. These associations remained significant even after adjusting for mean RT and demographic covariates. An implication of this finding is that the association between RTI and cognitive performance reflects more than artefact of the mean/SD confound in RT data. An important extension of the current study, however, was the demonstration that participants with RTI T-Scores of 65 or higher performed significantly more poorly on all five of the cognitive tests in the VLS battery. In addition, participants with RTI T-Scores of 45 or lower performed significantly better than individuals in the average range on the cognitive tests in the VLS battery. These are important findings for establishing the potential clinical utility of RTI because in clinical contexts, test scores are evaluated relative to mean performance of a normative group. However, it is noteworthy that the majority of results of the multinomial analyses were not

independent of Mean RT. As discussed in Study 1, Mean RT is associated with both RTI and cognitive performance. These results suggest that distributional associations between RTI and performance on cognitive tests in the VLS may be best explained by corresponding differences in Mean RT.

Similar findings were observed for analyses examining RTI T-Scores in relation to cognitive status. Across all four operationalizations of RTI and the four operationalizations of cognitive status, higher T-Scores were associated with increased risk of cognitive impairment. Consistent with the results reported in Study 1, these associations remained significant for the MA-MCI impairment classification after adjusting for Mean RT and demographic covariates, but the remaining three operationalizations of cognitive impairment were rendered nonsignificant. The residual ISD and the MSSD consistently yielded ROC curves that were suggestive of better classification accuracy than mean RT values. ROC curves for mean RT and all RTI scores all fell slightly below the accepted range for potential clinical utility (0.70; Hanley & MacNeil, 1982), and continuous RTI T-Scores had a poor 55% specificity for detecting cognitive impairment. However, the observation that RTI scores perform similarly or slightly better than mean RT for detecting cognitive impairment provides further support for the value of examining RTI values as standalone indices of CNS integrity.

Multinomial regression demonstrated that RTI T-scores in the high range are associated with a significant increase the odds of MA-MCI relative to T-Scores in the average range. Similarly, T-Scores in the low range significantly decreased the odds of impairment relative to the average range. These results suggest that RTI T-Scores conform to the distributional assumptions that underlie standardized assessment. An examination of stratum specific likelihood ratios further suggested that T-Scores obtained from all four operationalizations of

RTI have to potential to make a small contribution to clinical decision making. In particular, RTI scores were shown to demonstrate utility for detecting both "true positive" and "true negative" cases of cognitive impairment. Findings of a protective association between low RTI and cognitive performance has been reported previously, but most literature has focused on the pathological correlates of high RTI. The results of the current study provide further support for the utility of RTI T-Scores for identifying "consistent" examinees who are at reduced risk of diseases of aging.

In the assembly of the normative sample, both age and education were shown to significantly and independently predict RTI. As a result, all normative data were stratified by age and education, with 5-year age increments yielding slightly stronger associations with RTI relative to 10-year increments. Age effects of RTI have consistently been reported in the literature (Dykiert, Der, Starr, & Deary, 2012), and thus it was expected that age stratification would be necessary in order to avoid measurement bias. The finding of an association between RTI and educational attainment is consistent with prior research. The mechanisms linking education and RTI have not been directly examined in the literature, but the broader association between education and brain heath in old age is well documented (e.g., EClipSE Collaborative Members, 2010). Contrary to a body of literature documenting effects of sex on RTI (Dykiert, Der, Starr, & Deary, 2012b), the present study found no association between participant sex and RTI that was not better accounted for by age and education.

Several approaches to classifying cognitive impairment were examined in the current study in order to determine the relative sensitivity of measures of RTI to mild vs. stricter classifications of impairment. The MA-MCI approach was selected as our primary approach because it capitalizes on the longitudinal design of the VLS by requiring the presence of weak

cognitive performance across two waves of measurement in order to meet impairment criteria. This approach to classifying impairment has been shown in a separate sample to improve the reliability of distribution-based research classifications of MCI (Vandermorris, Hultsch, Hunter, MacDonald & Strauss, 2011). The results of this study found clear associations between RTI and MA-MCI that were independent of Mean RT. In contrast, adjustment for mean RT rendered nonsignificant associations between RTI T-Scores and the other three examined operationalizations of cognitive status.

The Victoria Longitudinal Study is a multifaceted study of clinical, cognitive and functional aspects of aging. It was not designed to study cognitive impairment or dementia, and VLS participants tend to be healthy, well-educated and cognitively intact. Thus, a clear limitation of the current study is the high-functioning nature of the study sample. Despite the implementation of the MA-MCI approach to classification of cognitive impairment, replication of the findings reported here is warranted in a sample with objective clinical impairment. Study 3 examines the replicability of findings obtained from the VLS in an independent clinical sample.

Study 3: Norm-Referenced Operationalizations of RTI: Validation in a Clinical Sample

Establishment of criterion validity is a critical step in the development and evaluation of clinical tools. However, in order for criterion validity studies to yield meaningful findings, it is necessary for the validation sample to closely resemble the population with which a clinical tool would be used (APA, 1999). Study 2, which examined the association between norm-referenced RTI measures and cognitive and physical functioning in participants of the Victoria Longitudinal Study (VLS), provided an important proof-of-principle by demonstrating that norm-referenced measures of RTI are sensitive to cognitive performance and cognitive status in a sample of healthy older adults who are functioning generally within the normal range of ability. However, participants of the VLS as a group are not representative of populations that are at risk of cognitive impairment, and those who do experience cognitive or functional decline tend to be more likely to withdraw from the study than healthy participants. The results of Study 2 are thus problematic for the purposes of establishing the clinical utility of norm-referenced RTI scores because the sensitivity of a measure to cognitive impairment cannot be inferred on the basis of its sensitivity to cognitive performance in healthy individuals. The purpose of the current study was to determine whether findings reported in Study 2, based on the generally healthy and high functioning VLS sample, can be replicated in a clinical sample.

PREVENT is a multivariate study of risk factors for Alzheimer's disease conducted at the University of Victoria. The purpose of the current study was to examine whether RT data from PREVENT could be standardized against the VLS normative sample assembled in Study 2 for the purpose of detecting individuals with impairment in physical or cognitive functioning. Specifically, this study examined linear relationships between RTI T-scores and performance on

tests of physical and cognitive function in participants of PREVENT. In addition, the sensitivity of RTI T-Scores to cognitive impairment was examined and the clinical utility of results was evaluated using Stratum Specific Likelihood Ratios.

Study 3 Methods

Participants

Participants of PREVENT were recruited via print and radio advertisements. Participants underwent neuropsychological assessment to determine their cognitive status and were classified as normal vs. impaired on the basis of their test scores. Objective measurements of physical functioning, including gait recording using the GaitRITE system, were also obtained from all participants. Eligibility criteria for PREVENT required that participants have no history of stroke, epilepsy, or other neurological disorder likely to affect cognition. Individuals who experienced a major depressive episode or any other psychiatric condition within a year prior to the screening interview for participation in PREVENT were also excluded.

Cognitive Assessment

Participants of PREVENT completed a 90-minute cognitive battery that was used to classify participants as having normal cognition, mild cognitive impairment, or probable Alzheimer's disease. The battery included scores from the following measures: Rey Auditory Verbal Learning Test (RAVLT), Benton Visual Retention Test (BVRT), Controlled Oral Word Association Test (COWAT), Animal Fluency, Trail Making Test (TMT) Part A and Part B, Clock Drawing Test, Mental Alternation Test (MAT), and WAIS-R subtests Digit-Symbol Substitution Test, Similarities (abbreviated), Block Design (abbreviated), and Digit Span (Wechsler, 1987).

Procedure: Two Masters-level clinicians conducted a clinical interview with each participant regarding their impression of their current cognitive functioning and any history of illness or injury that could impact cognitive performance. Where necessary, information obtained from participants was corroborated by interviewing an informant close to the participant.

Participants completed a cognitive assessment battery and test scores were standardized against appropriate normative data (normative data from the Canadian Study of Health and Aging for all measures apart from the Trail Making Test; data from the Mayo Older American Normative Sample for the Trail Making Test). Participants with performance falling 1 standard deviation below the mean on one or more test of memory were classified as having Mild Cognitive Impairment (MCI). Those with memory performance falling 2 or more standard deviations below average and impairment in at least one other cognitive domain were diagnosed with probable Alzheimer's disease provided information from the clinical interview corroborated this diagnosis. The PREVENT diagnosis of Alzheimer's disease corresponds to NINCDS-ADRDA criteria for probable Alzheimer's disease.

Assessment of Physical Functioning

Participants of PREVENT completed several objective tests of physical functioning, including grip strength, two measures of peak flow, and a total of 6 blood pressure measurements divided into two occasions during the testing appointment (4 arm measurements, 2 ankle measurements). Gait was measured in PREVENT using GAITRite, an instrumented walkway with embedded pressure sensors and 20 feet of recording surface. Gait parameters obtained by GAITRite include, but are not limited to, gait speed (cm/second), stride length, cadence, and stride-time variability. These gait parameters have been shown to have good reliability and validity (Branch, Perera, Studenski, et al., 2008), and are among the most widely used parameters in gait research. Of the above parameters, the current study examines gait speed and stride-time variability based on their previously documented association with cognitive status in older adults (Martin et al., 2013).

Procedure: Participants of PREVENT completed three gait conditions, each consisting of five trials. The first condition involved walking freely across the mat. The second and third conditions were included to examine the effects of cognitive load on gait parameters, with the second condition requiring participants to spell 5-7-letter words backwards while walking, and the third condition requiring participants to count backwards by sevens during the gait recording. Prior to all trials participants were instructed to walk at a normal rate. The current study examines gait parameters derived from the free walk and the dual-task walk in which participants completed the serial sevens task.

Reaction Time Measures

Reaction time measures included in PREVENT were the SRT and Lexical decision tasks described in detail in Study 1 (pg. 17-18).

Classification of Participants

Contrary to the primary cognitive outcome of the PREVENT study, the current analyses classified cognitive functioning using multi-domain cognitive impairment classification followed by Strauss and colleagues (2006) and Dixon and colleagues (2007). Participants were classified as impaired based on the presence of performance falling 1.5 standard deviations or more below the mean on at least one measure of memory and at least one measure of an additional domain of cognitive functioning (e.g., executive functioning, visualspatial ability). Physical functioning was evaluated using continuous scores obtained from the objective measures of physical functioning that were included in the PREVENT study battery.

Standardization of RTI Data

RTI data from participants of PREVENT were standardized against normative data obtained from Sample 2 of the VLS following procedures described in Study 2. Consistent with

the methods described in Study 2, RTI T-scores were examined as continuous variables in bivariate analyses and in the binomial logistic regression analyses with cognitive status as the outcome. RTI T-scores were then examined categorically in multinomial logistic regression analyses with RTI stratum as the outcome and continuous measures of physical and cognitive function as predictors. Consistent with Study 2, categorical operationalization of RTI T-scores followed psychometric distribution-based classifications of impairment. T-scores of 65 and higher were coded as "2", those ranging from 45-64 were coded as "1", those below 45 were coded as "0".

Statistical Analyses

Data Preparation: Preparation of RT data proceeded as described in Study 1. RTI T-Scores standardized against the VLS normative sample were examined in all analyses. Demographic characteristics of the sample were examined by cognitive status and RTI T-Score stratum to determine whether these groups differed systematically according to any relevant demographic dimension.

Linear Regression Analyses: Linear regression was used to examine associations between continuous RTI T-Scores and continuous cognitive and physical outcomes. These associations were examined first in univariate models with RTI T-Scores as the sole predictor, and subsequently along with mean RT in forward selection linear regression models adjusted for age, sex and education. Analyses of gait further adjusted for leg length and gait velocity. The forward regression analyses allow for the determination of the relative utility of RTI and mean RT for predicting physical and cognitive performance. To account for multiple comparisons, only *p*-values < 0.01 were interpreted as statistically significant.

Binomial Logistic Regression Analyses: Binomial logistic regression was used to examine the association between continuous RTI T-Scores and cognitive status in the PREVENT sample. As with the linear regression analyses, the predictive value of RTI T-Scores for detecting cognitive impairment was examined first in a series of univariate logistic regression models, and subsequently in forward regression models along with mean RT. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Multinomial Logistic Regression Analyses: Multinomial logistic regression was used to determine whether the Study 2 findings of higher proportion of cognitive impairment in examinees with RTI scores in the "high" range, and a lower proportion of cognitive impairment in examinees with RTI scores in the "low" range could be replicated in a sample with objective cognitive impairment. Consistent with Study 2, T-Score values of 65 or higher were classified as "high" RTI, and values below 45 were classified as "low" RTI. The classification of values of 65 and higher as "high" RTI is in keeping with conventional assessment cutoffs that classify all values falling 1.5 SD beyond the mean as clinically elevated. Multinomial regression models were also used to examine performance on objective tests of physical and cognitive functioning of examinees in "high" and "low" RTI strata relative to those in the average range. To account for multiple comparisons, only p-values < 0.01 were interpreted as statistically significant.

Stratum-Specific Likelihood Ratios: As described in Study 2, Stratum-specific likelihood ratios (SSLRs) were examined in the PREVENT sample to determine whether differences in the prevalence of cognitive impairment as a function of RTI strata were of a sufficient magnitude to implicate clinical utility of measures of RTI (Jaeschke, Guyatt & Sackett, 1994).

Study 3 Results

Demographic characteristics

Demographic characteristics of the study sample are presented by cognitive status in Table 19. Participants in the healthy sample had a significantly higher level of educational attainment than the cognitively impaired sample (15.31 years vs. 13.60 years). In addition, the healthy sample significantly outperformed the impaired sample on the Modified Mini Mental Status exam, mean RT on the Lexical Decision test, and on neuropsychological measures of processing speed and delayed verbal recall. Groups did not differ significantly in age, sex, or performance on objective measures of physical functioning.

Table 19. Demographic characteristics of the PREVENT sample.

	Healthy	Impaired	*F* (*df*), **p*
N	54	26	
Age	73.27 (11.71)	76.12 (6.17)	1.34 (1), *NS*
Sex, %F	48%	42%	0.24 (1), *NS*
Education	**15.31 (3.23)**	**13.60 (3.12)**	**4.86 (1)***
Peak flow	374.59 (95.38)	330.79 (97.59)	2.97 (1), *NS*
Grip strength	33.68 (10.01)	28.39 (9.38)	3.69 (1), *NS*
3MS	**94.88 (4.84)**	**85.84 (12.66)**	**21.29 (1)****
RAVLT 7	**7.35 (3.47)**	**3.21 (3.77)**	**24.22 (1)****
Digit symbol	**45.21 (12.61)**	**32.86 (12.95)**	**14.23(1)****
Lexical	**1149.94 (346.90)**	**1483.03 (761.59)**	**8.09 (1)****

Note. RAVLT 7 = Trial7 of the Rey Auditory Verbal Learning Test.

Standardization of PREVENT RTI data against the VLS normative sample

Bivariate associations among the mean RT and the four operationalizations of RTI in PREVENT relative to the normative sample are presented in Table 20a. Associations with mean RT and correlations among the four operationalizations of RTI were virtually identical in PREVENT and the VLS normative sample.

Table 20a. Correlation matrices for Lexical RTI Scores in VLS Sample 2 and PREVENT

	VLS Normative Sample Lexical RTI Scores				PREVENT Lexical RTI Scores			
	Raw ISD	Res ISD	CoV	MSSD	Raw ISD	Res ISD	CoV	MSSD
Lexical								
Mean	**.81**	**.81**	**.50**	**.80**	**.76**	**.76**	**.41**	**.72**
Raw ISD	**1.00**	**.99**	**.89**	**.99**	**1.00**	**.99**	**.89**	**.97**
Res ISD		**1.00**	**.89**	**.99**		**1.00**	**.89**	**.98**
CoV			**1.00**	**.88**			**1.00**	**.88**
MSSD				**1.00**				**1.00**

Note: ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; RTI = reaction time inconsistency.

Distributional characteristics of raw and standardized PREVENT RTI scores are presented in Table 20b. As described in the Methods (pg. 103), a categorical score was computed from RTI T-Scores by classifying scores as falling within "Low", "Average" and "High" ranges. Characteristics of participants falling within each of these T-Score Strata are presented in Table 21. There were no significant differences in age, sex or objective physical functioning of participants across low, average and high RTI strata for any of the four RTI computations. The prevalence of impairment was significantly lower in participants with RTI scores in the "low" range relative to the other groups. There was also a strong association between mean RT and RTI strata across all four computations of RTI, such that individuals in the "low" RTI T-Score range had faster mean RT and those in the "high" RTI T-Score range has slower mean RT.

RTI T-Scores in relation to physical functioning

Linear associations between RTI T-Scores and performance on objective measures of physical functioning in the PREVENT sample are reported in Table 22. In univariate models, significant associations between RTI T-Scores and variability in gait performance were observed for both the free gait condition and the dual-task gait condition. Specifically, T-Scores associated with the residual ISD and the raw ISD significantly predicted variability in dual-task gait conditions. The CoV and MSSD did not yield any association with gait variability. No association between RTI T-Scores and gait variability emerged from the demographically adjusted forward regression models. Rather, mean RT emerged as the sole RT predictor of variability in gait.

Table 20b. Distributional characteristics of raw and standardized RTI scores in PREVENT.

	Healthy						Impaired					
RTI Score	Mean	SD	Skewness	SE Skew	Kurt	SE Kurt	Mean	SD	Skewness	SE Skew	Kurt	SE Kurt
Unstandardized												
Raw ISD	2.410	0.268	0.339	0.327	-0.581	0.644	2.63	0.212	1.279	0.456	1.547	0.887
Residual ISD	-0.407	0.260	0.477	0.327	-0.508	0.644	-0.190	0.212	1.302	0.456	1.589	0.887
CoV	-0.632	0.187	0.001	0.327	-0.607	0.644	-0.514	0.098	0.706	0.456	1.068	0.887
MSSD	2.523	0.271	0.361	0.327	-0.355	0.644	2.743	0.200	1.279	0.456	1.547	0.887
T-Score												
Raw ISD	46.20	13.30	0.420	0.327	0.041	0.644	55.43	11.34	1.690	0.456	3.470	0.887
Residual ISD	44.90	12.70	0.551	0.327	-0.135	0.644	55.39	11.30	1.170	0.456	1.380	0.887
CoV	43.79	15.72	-0.047	0.327	-0.202	0.644	53.34	10.96	1.655	0.456	3.112	0.887
MSSD	45.52	12.96	0.547	0.327	0.341	0.644	54.27	9.48	1.250	0.456	2.060	0.887

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; SD = standard deviation; SE = standard error.

Table 21. Participant characteristics by RTI strata.

	Low RTI	Average RTI	High RTI	*F* (df), *p*
Raw ISD Strata				
Cell Size	30	40	9	
Age	71.60 (15.49)	73.52 (5.40)	74.38 (5.42)	0.590 (2) *NS*
Sex	42%	44%	78%	1.973 (2) *NS*
Education	15.70 (3.47)	14.39 (3.47)	13.75 (2.60)	1.887 (2) *NS*
Cognitive Status	**10%**	**50%**	**33%**	**7.091 (2)****
Systolic BP	135.49 (21.12)	133.13 (12.74)	137.37 (17.40)	0.300 (2) *NS*
Grip Strength	30.78 (8.92)	31.32 (10.21)	36.75 (11.58)	0.959 (2) *NS*
Peak Flow	374.20 (86.24)	352.32 (100.33)	375.00 (126.92	0.300 (2) *NS*
Mean Lexical	**973.54 (209.61)**	**1222.60 (252.46)**	**2188.35 (976.48)**	**30.198 (2)****
Res ISD Strata				
Cell Size	31	39	9	
Age	72.21 (14.88)	72.96 (4.90)	74.88 (5.87)	0.791 (2) *NS*
Sex	43%	43%	78%	1.964 (2) *NS*
Education	15.37 (3.46)	14.85 (3.46)	12.75 (2.71)	2.377 (2) *NS*
Cognitive Status	**12%**	**46%**	**44%**	**5.046 (2)****
Systolic BP	135.79 (21.71)	132.68 (10.14)	137.00 (17.50)	0.008 (2) *NS*
Grip Strength	31.25 (8.67)	31.09 (10.73)	36.13 (11.30)	1.172 (2) *NS*
Peak Flow	374.94 (82.65)	345.65 (104.50)	387.50 (122.19)	0.433 (2) *NS*
Mean Lexical	**1007.56 (226.48)**	**1198.24 (222.14)**	**2247.91 (942.22)**	**40.659 (2)****
CoV Strata				

Cell size	34	39	8	
Age	74.59 (7.16)	71.13 (13.66)	73.38 (5.18)	0.390 (2) *NS*
Sex	44%	46%	63%	0.436 (2) *NS*
Education	**15.72 (3.49)**	**14.66 (3.38)**	**12.38 (2.20)**	**3.572 (2)***
Cognitive Status	**12%**	**50%**	**38%**	**6.515 (2)****
Systolic BP	135.73 (19.54)	131.52 (13.87)	141.81 (16.00)	0.435 (2) *NS*
Grip Strength	31.02 (9.20)	31.87 (10.09)	34.09 (12.49)	0.280 (2) *NS*
Peak Flow	385.12 (84.53)	339.60 (95.57)	378.75 (135.47)	0.499 (2) *NS*
Mean Lexical	**1037.66 (246.99)**	**1334.54 (653.21)**	**1628.39 (599.53)**	**5.868 (2)****
MSSD Strata				
Cell size	30	43	8	
Age	72.85 (15.04)	74.23 (6.39)	73.25 (5.26)	0.192 (2) *NS*
Sex	43%	44%	75%	1.405 (2) *NS*
Education	15.48 (3.30)	14.45 (3.33)	13.88 (2.47)	1.226 (2) *NS*
Cognitive Status	**10%**	**48%**	**38%**	**6.069 (2)****
Systolic BP	138.88 (21.52)	135.68 (15.17)	137.62 (17.55)	0.271 (2) *NS*
Grip Strength	32.26 (9.64)	30.70 (9.46)	35.94 (11.18)	1.068 (2) *NS*
Peak Flow	385.68 (82.71)	339.54 (109.66)	397.50 (135.49)	1.548 (2) *NS*
Mean Lexical	**975.67 (218.40)**	**1279.29 (338.29)**	**2170.12 (985.12)**	**27.20 (2)****

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean square successive difference; All values presented as mean (standard deviation) unless noted otherwise.

Table 22. Lexical RTI T-Scores in relation to objective measures of physical function.

	Unadjusted			Adjusted
	B (95% CI)	*p*	R2	B (95% CI)
Raw ISD T-Score				
Systolic BP	-0.071 (-0.369, 0.226)	.635	.003	NS
Diastolic BP	0.037 (-0.154, 0.227	.703	.002	NS
Pulse	0.016 (-0.184, 0.216)	.876	.000	NS
Grip Strength	0.065 (-0.104, 0.233)	.448	.007	NS
Peak Flow	0.014 (-1.809, 1.837)	.988	.000	NS
Gait Velocity-Free	-0.369 (-1.009, 0.271)	.254	.005	NS
Gait Velocity-Load	-0.372 (-1.031, 0.288)	.265	.004	NS
Stride Variability-Free	0.003 (0.000, 0.007)	.060	.037	NS
Stride Variability-Load	**0.002 (0.001, 0.003)**	**.001**	**.145**	NS
Res ISD T-Score				
Systolic BP	0.023 ('-0.280, 0.326)	.881	.000	NS
Diastolic BP	0.027 (-0.166, 0.221)	.779	.001	NS
Pulse	-0.019 (-0.222, 0.184)	.853	.000	NS
Grip Strength	0.023 (-0.149, 0.195)	.788	.001	NS
Peak Flow	-0.271 (-2.108, 1.567)	.988	.000	NS
Gait Velocity-Free	-0.215 (-1.246, 0.049)	.070	.033	NS
Gait Velocity-Load	0.018 (-1.178, 0.168)	.139	.018	NS
Stride Variability-Free	0.004 (0.000, 0.008)	.026	.056	NS
Stride Variability-Load	**0.009 (0.003, 0.016)**	**.000**	**.111**	NS

CoV T-Score				
Systolic BP	0.019 (-0.250, 0.289)	.888	.000	NS
Diastolic BP	0.058 (-0.114, 0.230)	.506	.006	NS
Pulse	-0.037 (-0.218, 0.143)	.683	.002	NS
Grip Strength	0.043 (-0.108, 0.194)	.573	.004	NS
Peak Flow	-0.180 (-1.798, 1.437)	.825	.001	NS
Gait Velocity-Free	-0.224 (-0.807, 0.359)	.446	.006	NS
Gait Velocity-Load	-0.090 (-0.695, 0.514)	.766	.001	NS
Stride Variability-Free	0.000 (-0.003, 0.003)	.925	.000	NS
Stride Variability-Load	0.004 (-0.002, 0.010)	.186	.011	NS
MSSD T-Score				
Systolic BP	-0.100 (-0.422, 0.223)	.540	.005	NS
Diastolic BP	0.051 (-0.142, 0.243)	.602	.004	NS
Pulse	-0.016 (-0.218, 0.187)	.878	.000	NS
Grip Strength	0.038 (-0.139, 0.215)	.672	.002	NS
Peak Flow	-0.183 (-2.083, 1.717)	.848	.000	NS
Gait Velocity-Free	-0.342 (-1.038, 0.355	.331	.001	NS
Gait Velocity-Load	-0.391 (-1.111, 0.329)	.282	.003	NS
Stride Variability-Free	0.003 (-0.001, 0.007)	.104	.024	NS
Stride Variability-Load	0.008 (0.001, 0.014)	.030	.057	NS

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant.

RTI T-Scores in relation to cognitive performance

Linear associations between RTI T-Scores and performance on the PREVENT neuropsychological test battery are reported in Table 23. In univariate models, RTI T-Scores significantly predicted performance on all examined neuropsychological test scores except for Digit Span Forward and Digit Span Backward. R-Square estimates suggested that RTI T-Scores most strongly predicted performance on RAVLT Trial 7, BVRT and Digit Symbol. In the demographically adjusted forward regression models, RTI T-Scores significantly predicted performance on the majority of examined tests. In particular, RTI T-Score operationalizations emerged from forward regression analyses as predictors of verbal recall, phonemic fluency, semantic fluency, similarities and digit symbol. Mean RT did not yield any association with these tests independent of RTI T-Scores. In contrast, Mean RT emerged from all four forward analyses as the sole RT predictor of verbal learning, digit span and TMT A and B.

Multinomial logistic regression analyses were used to compare the cognitive performance of participants with RTI T-Scores in the "low" and "high" ranges to those with T-Scores in the average range. Results are presented in Table 24. Results did not support the presence of any significant differences in cognitive performance between participants with RTI T-Scores in the average and high ranges. In contrast, select significant associations were observed between raw ISD, residual ISD, and CoV T-Scores in the low range and performance on neuropsychological tests in the PREVENT battery including verbal recall, semantic fluency, and digit symbol. The pattern of significant associations was not consistent across RTI T-Scores. Despite the limited significant findings, it is noteworthy that nearly all tests in the battery shared associations with low RTI T-Scores that would meet conventional (e.g., $p < 0.05$) standards for statistical significance.

Table 23. RTI T-Scores as predictors of PREVENT neuropsychological test scores.

	Unadjusted			Adjusted		
	B (95% CI)	*p*	R2	B (95% CI)	*p*	R2 Adj
Raw ISD T-Score						
RAVLT 1-5	**-0.405 (-0.530, -0.157)**	**.000**	**.152**	NS		
RAVLT Trial 6	**-0.450 (-0.190, -0.067)**	**.000**	**.191**	**-0.109 (-0.172, -0.046)**	**.001**	**.231**
RAVLT Trial 7	**-0.410 (-0.190, -0.057)**	**.000**	**.156**	**-0.103 (-0.171, -0.036)**	**.003**	**.194**
BVRT	**-0.453 (-0.113, -0.040)**	**.000**	**.194**	**-0.002 (-0.003, -0.001)**	**.000**	**.226**
Block Design	**-0.250 (-0.205, -0.006)**	**.000**	**.048**	**-0.340 (-0.578, -0.103)**	**.007**	**.413**
Digit Symbol	**-0.536 ('0.741, -0.325)**	**.000**	**.276**	**-0.493 (-0.704, -0.282)**	**.000**	**.325**
TMT A	**0.333 (-0.212, 1.111)**	**.000**	**.098**	NS		
TMT B	**0.335 (0.775, 3.892)**	**.004**	**.099**	NS		
Digit Span F	-0.231 (-0.079, 0.000)	.051	.040	NS		
Digit Span B	-0.185 (-0.059, 0.007)	.126	.020	NS		
Phonemic Fluency	**-0.402 (-0.586, -0.171)**	**.001**	**.149**	**-0.355 (-0.572, -0.137)**	**.002**	**.157**
Semantic Fluency	**-0.378 (-0.264, -0.067)**	**.001**	**.130**	**-0.173 (-0.274, -0.072)**	**.001**	**.171**
Similarities	**-0.318 (-0.104, -0.017)**	**.007**	**.088**	NS		
Res ISD T-Score						
RAVLT 1-5	**-0.427 (-0.562, -0.183)**	**.000**	**.170**	NS		
RAVLT Trial 6	**-0.135 (-0.198, -0.073)**	**.000**	**.200**	**-0.113 (-0.178, -0.047)**	**.001**	**.229**
RAVLT Trial 7	**-0.135 ('-0.202, -0.067)**	**.000**	**.176**	**-0.111 (-0.181, -0.040)**	**.003**	**.200**
BVRT	**-0.076 (-0.114, -0.039)**	**.000**	**.180**	NS		

Block Design	-0.110 (-0.213, -0.008)	.035	.051	-0.123 (-0.229, -0.018)	.022	.105
Digit Symbol	**-0.563 (-0.774, -0.351)**	**.000**	**.292**	**-0.505 (-0.726, -0.284)**	**.000**	**.318**
TMT A	**0.765 (0.310, 1.220)**	**.001**	**.128**	NS		
TMT B	**2.679 (1.091, 4.268)**	**.001**	**.129**	NS		
Digit Span F	-0.044 (-0.085, -0.003)	.035	.048	-0.047 (-0.091, -0.002)	.039	.013
Digit Span B	-0.029 (-0.063, 0.005)	.090	.028	NS		
Phonemic Fluency	**-0.445 (-0.652, -0.237)**	**.000**	**.198**	**-0.410 (-0.632, -0.188)**	**.000**	**.188**
Semantic Fluency	**-0.191 (-0.290, -0.092)**	**.000**	**.167**	**-0.190 (-0.294, -0.086)**	**.001**	**.188**
Similarities	**-0.076 (-0.119, -0.033)**	**.001**	**.139**	**-0.066 (-0.112, -0.019)**	**.006**	**.135**
CoV T-Score						
RAVLT 1-5	**-0.280 (-0.446, -0.113)**	**.001**	**.128**	NS		
RAVLT Trial 6	**-0.098 (-0.154, -0.043)**	**.001**	**.140**	**-0.085 (-0.140, -0.029)**	**.003**	**.203**
RAVLT Trial 7	**-0.103 (-0.162, -0.045)**	**.001**	**.139**	**-0.0 (-0.149, -0.031)**	**.003**	**.193**
BVRT	**-0.060 (-0.093, -0.027)**	**.001**	**.149**	-0.038 (-0.072, -0.004)	.028	.323
Block Design	0.096 (-0.184, -0.008)	.033	.052	-0.100 (-0.188, -0.012)	.027	.101
Digit Symbol	**-0.394 (-0.588, -0.201)**	**.000**	**.190**	**-0.261 (-0.458, -0.064)**	**.010**	**.333**
TMT A	0.317 (-0.096, 0.730)	.130	.190	NS		
TMT B	1.447 (0.022, 2.872)	.047	.042	NS		
Digit Span F	-0.042 (-0.076, -0.008)	.017	.065	-0.043 (-0.079, -0.007)	.021	.029
Digit Span B	-0.027 (-0.056, 0.002)	.063	.036	NS		
Phonemic Fluency	**-0.328 (-0.511, -0.145)**	**.001**	**.144**	**-0.301 (-0.491, -0.112)**	**.002**	**.152**
Semantic Fluency	**-0.130 (-0.218, -0.041)**	**.005**	**.098**	NS		
Similarities	-0.047 (-0.086, -0.009)	.016	.067	NS		

MSSD T-Score

RAVLT 1-5	**-0.326 (-0.529, -0.122)**	**0.002**	0.116	NS		
RAVLT Trial 6	**-0.126 (-0.193, -0.058)**	**0.000**	0.156	**-0.106 (-0.173, -0.038)**	**.003**	**.208**
RAVLT Trial 7	**-0.121 (-0.193, -0.049)**	**0.001**	0.127	**-0.101 (-0.174, -0.028)**	**.007**	**.176**
BVRT	**-0.071 (-0.111, -0.031)**	**0.001**	0.142	NS		
Block Design	**-0.127 (-0.234, -0.021)**	**0.019**	**0.065**	**-0.147 (-0.252, -0.041)**	**.007**	**.133**
Digit Symbol	**-0.533 (-0.762, -0.304)**	**0.000**	**0.238**	**-0.493 (-0.721, -0.265)**	**.000**	**.300**
TMT A	**0.740 (0.261, 1.219)**	**0.003**	**0.108**	NS		
TMT B	**2.384 (0.694, 4.074)**	**0.006**	**0.090**	NS		
Digit Span F	-0.047 (-0.089, -0.004)	0.031	0.051	-0.049 (-0.094, -0.004)	.032	.018
Digit Span B	-0.039 (-0.074, -0.005)	0.027	0.056	-0.044 (-0.080, -0.008)	.017	.068
Phonemic Fluency	**-0.378 (-0.603, -0.153**	**0.001**	**0.128**	**-0.353 (-0.586, -0.121)**	**.003**	**.142**
Semantic Fluency	**-0.174 (-0.280, -0.069)**	**0.002**	**0.125**	**-0.181 (-0.288, -0.074)**	**.001**	**.169**
Similarities	-0.053 (-0.100, -0.006)	0.027	0.055	NS		

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; TMT = trail making test; RAVLT = Rey Auditory Verbal Learning Test; BVRT = Benton Visual Retention Test; * = p-values < 0.05; ** = p-values < 0.01.

Table 24. Multinomial regression analysis of cognitive performance by RTI strata.

	Low	Average	High
Raw ISD Strata			
RAVLT 1-5	1.050 (0.998, 1.105) *NS*		0.948 (0.888, 1.012) *NS*
RAVLT Trial 6	1.177 (1.019, 1.361)*		0.779 (0.598, 1.014) *NS*
RAVLT Trial 7	**1.212 (1.052, 1.398)****		0.902 (0.733, 1.110) *NS*
BVRT	1.440 (1.013, 2.047)*		1.000 (0.752, 1.330) *NS*
Block Design	1.113 (1.006, 1.231)*		1.108 (0.971, 1.266) *NS*
Digit Symbol	1.085 (1.028, 1.145)*		0.961 (0.902, 1.025) *NS*
TMT A	0.951 (0.907, 0.997)*		1.010 (0.991, 1.030) *NS*
TMT B	0.990 (0.980, 0.999)*		1.001 (0.995, 1.008) *NS*
Digit Span F	1.392 (1.076, 1.801)*		1.012 (0.731, 1.401) *NS*
Digit Span B	1.198 (0.905, 1.586) *NS*		0.847 (0.565, 1.268) *NS*
Phonemic Fluency	1.046 (1.000, 1.093*		0.980 (0.924, 1.040) *NS*
Semantic Fluency	1.131 (1.025, 1.249)*		0.954 (0.833, 1.092) *NS*
Similarities	1.082 (0.860, 1.361) *NS*		0.857 (0.677, 1.085) *NS*
Residual ISD Strata			
RAVLT 1-5	1.029 (0.983, 1.078) *NS*		0.948 (0.883, 1.018) *NS*
RAVLT Trial 6	1.133 (0.989, 1.297) *NS*		0.775 (0.579, 1.039) *NS*
RAVLT Trial 7	1.146 (1.007, 1.303)*		0.914 (0.732, 1.142) *NS*
BVRT	1.314 (0.971, 1.779) *NS*		0.926 (0.697, 1.231) *NS*
Block Design	1.081 (0.984, 1.187) *NS*		1.101 (0.956, 1.267) *NS*
Digit Symbol	**1.070 (1.019, 1.124)****		0.959 (0.895, 1.028) *NS*

TMT A	0.975 (0.942, 1.010) *NS*	1.010 (0.989, 1.030) *NS*
TMT B	0.993 (0.985, 1.000)*	1.002 (0.994, 1.009) *NS*
Digit Span F	1.312 (1.032, 1.669)*	0.996 (0.697, 1.422) *NS*
Digit Span B	1.180 (0.898, 1.550) *NS*	0.763 (0.481, 1.210) *NS*
Phonemic Fluency	1.048 (1.004, 1.095)*	0.961 (0.898, 1.030) *NS*
Semantic Fluency	1.111 (1.012, 1.219)*	0.969 (0.837, 1.123) *NS*
Similarities	1.091 (0.865, 1.375) *NS*	0.779 (0.603, 1.006) *NS*
CoV Strata		
RAVLT 1-5	1.053 (1.003, 1.106)*	0.991 (0.924, 1.062) *NS*
RAVLT Trial 6	**1.214 (1.052, 1.401)****	0.964 (0.765, 1.215) *NS*
RAVLT Trial 7	**1.210 (1.055, 1.387)****	1.028 (0.837, 1.263) *NS*
BVRT	1.446 (1.027, 2.036)*	0.976 (0.721, 1.321) *NS*
Block Design	1.087 (0.989, 1.196) *NS*	1.098 *0.953, 1.265) *NS*
Digit Symbol	1.082 (1.028, 1.139)*	0.995 (0.930, 1.064) *NS*
TMT A	0.964 (0.926, 1.003) *NS*	0.986 (0.946, 1.029) *NS*
TMT B	0.990 (0.982, 0.999)*	0.997 (0.988, 1.006) *NS*
Digit Span F	1.349 (1.054, 1.727)*	1.067 (0.743, 1.531) *NS*
Digit Span B	1.251 (0.947, 1.651) *NS*	1.041 (0.674, 1.608) *NS*
Phonemic Fluency	1.057 *0.925, 1.055)*	0.988 (0.925, 1.055) *NS*
Semantic Fluency	**1.141 (1.034, 1.258)****	1.065 (0.923, 1.230) *NS*
Similarities	1.156 (0.925, 1.445) *NS*	1.066 (0.773, 1.471) *NS*
MSSD Strata		
RAVLT 1-5	1.027 (0.980, 1.077) *NS*	0.940 (0.881, 1.003) *NS*

RAVLT Trial 6	1.182 (1.024, 1.364)*	0.818 (0.642, 1.042) *NS*
RAVLT Trial 7	1.184 *1.033, 1.357)*	0.878 (0.714, 1.082) *NS*
BVRT	1.523 (1.061, 2.186)*	0.903 (0.702, 1.161) *NS*
Block Design	1.135 (1.023, 1.258)*	1.039 (0.903, 1.196) *NS*
Digit Symbol	1.081 (1.025, 1.139)*	0.957 (0.896, 1.022) *NS*
TMT A	0.964 (0.925, 1.004) *NS*	1.013 (0.993, 1.033) *NS*
TMT B	0.990 (0.981, 0.999)*	1.002 (0.996, 1.009) *NS*
Digit Span F	1.273 (1.003, 1.615)*	0.992 (0.728, 1.351) *NS*
Digit Span B	1.224 (0.926, 1.618) *NS*	0.897 (0.608, 1.325) *NS*
Phonemic Fluency	1.053 (1.008, 1.101)*	0.960 (0.899, 1.026) *NS*
Semantic Fluency	1.116 (1.015, 1.227)*	0.930 (0.804, 1.077) *NS*
Similarities	1.227 (0.958, 1.571) *NS*	0.885 (0.694, 1.127) *NS*

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = non-significant; TMT = trail making test; RAVLT = Rey Auditory Verbal Learning Test; BVRT = Benton Visual Retention Test; * = p-values < 0.05; ** = p-values < 0.01.

RTI T-Scores in relation to cognitive status

Binomial logistic regression models were used to examine the sensitivity of Lexical RTI T-Scores to cognitive status. These associations were examined first using univariate models with RTI T-Scores as the sole predictor of cognitive status, and subsequently using adjusted models with age, sex, education and mean RT included as covariates. Results are presented in Table 25. In the univariate models, all four RTI T-Scores significantly predicted cognitive status,

with the residual ISD T-Scores accounting for the most variance in cognitive status. In the adjusted forward regression models, RTI T-Scores for the raw ISD, the Residual ISD and the MSSD remained significant predictors of cognitive status. Mean RT did not significantly predict cognitive status independent of these RTI T-Scores. In contrast, CoV T-Scores did not significantly predict cognitive status independent of mean RT. The Residual ISD T-Score had 73% sensitivity and 56% specificity when the cutoff value of the predicted probability of impairment was 0.34.

Table 25. Logistic analyses of RTI T-Scores as predictors of cognitive status

	Unadjusted			Forward Selection Model		
	B (95% CI)	*p*	R2 Adj	B (95% CI)	*p*	R2 Adj
Raw ISD	**1.059 (1.016, 1.103)**	**.007**	**.104**	**1.059 (1.016, 1.104)**	**0.007**	**.106**
Res ISD	**1.070 (1.025, 1.118)**	**.002**	**.137**	**1.071 (1.025, 1.119)**	**0.002**	**.141**
CoV	**1.050 (1.011, 1.089)**	**.002**	**.092**	NS		
MSSD	**1.063 (1.018, 1.111)**	**.006**	**.106**	**1.062 (1.018, 1.112)**	**0.006**	**.110**

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant;

Multinomial regression was used to examine whether the prevalence of cognitive impairment differed significantly in participants with “high” or “low” RTI T-Scores relative to those with T-Scores in the average range. Results are presented in Table 26. There were no significant differences in the prevalence of cognitive impairment in examinees with RTI T-

Scores in the high range vs. the average range. However, the prevalence of cognitive impairment was significantly lower in examinees with RTI scores in the low range relative to the average range. This association was observed for T-Scores associated with the raw ISD, the residual ISD and the MSSD. As presented in Table 27, SSLRs indicated that the presence of raw ISD, residual ISD and MSSD T-Scores in the low range can make a small but potentially clinically meaningful contribution to the process of ruling out the presence of cognitive impairment (Jaeschke, Guyatt & Sackett, 1994). No differences in the prevalence of cognitive impairment were observed across the CoV RTI strata and so SSLRs were not examined for this operationalization of RTI.

Table 26. RTI T-Score Strata: Association with Cognitive Status

	Low	Average	High
Raw ISD	**0.111 (0.029, 0.426)****	`	0.500 (0.110, 2.282) *NS*
Res ISD	**0.173 (0.051, 0.588)****		0.933 (0.217, 4.010) *NS*
CoV	**0.138 (0.041, 0.469)****		0.600 (0.125, 2.873) *NS*
MSSD	**0.127 (0.033, 0.485)****		0.660 (0.139, 3.123) *NS*

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; NS = nonsignificant; * = p-values < 0.05; ** = p-values < 0.01.

Table 27. Stratum-specific likelihood ratios associated with RTI T-Score strata.

	Cognitive Status				
Raw ISD Cutoff	Impaired	Healthy	%Impaired	% Healthy	SSLR
Low	3	27	0.115	0.509	**0.226**
Average	20	20	0.769	0.377	Ref
High	3	6	0.115	0.113	1.018
Res ISD Cutoff	Impaired	Healthy	%Impaired	% Healthy	SSLR
Low	4	27	0.154	0.509	**0.303**
Average	18	21	0.692	0.396	Ref
High	4	5	0.154	0.094	1.638
MSSD Cutoff	Impaired	Healthy	%Impaired	% Healthy	SSLR
Low	3	26	0.115	0.491	**0.234**
Average	20	22	0.769	0.415	Ref
High	3	5	0.115	0.094	1.223

Note. RTI = reaction time inconsistency; ISD = intraindividual standard deviation; CoV = coefficient of variation; MSSD = mean squared successive difference; SSLR = stratum specific likelihood ratio.

Study 3 Discussion

The current study sought to determine whether the results of Study 2 could be replicated in the clinical research sample of PREVENT. This extension of Study 2 was important from a test validation perspective because, in order to demonstrate the clinical utility of a test, criterion validity must be demonstrated in a sample that resembles the population in which clinical use is intended. The results of this study suggest that RTI T-Scores share similar associations with cognitive performance in the PREVENT sample as were observed in the VLS. In addition, this study replicated the finding of clinically significant effects of low RTI T-Scores for ruling out the presence of cognitive impairment. The finding observed in the VLS of clinical utility of high RTI T-Scores for positively predicting impairment, however, was not replicated in PREVENT.

Consistent with Study 2, RTI T-Scores independently predicted performance on the PREVENT neuropsychological test battery. In addition, forward regression analyses indicated that RTI T-Scores accounted for more variability in performance on several tests in the PREVENT battery than mean RT. Associations were strongest with tests of delayed verbal recall, verbal fluency and digit symbol. The finding of an association between RTI T-Scores and performance on Digit Symbol is inconsistent with the results observed from VLS, where Digit Symbol was not associated with RTI independent of mean RT. A potential explanation for this may be the older average age and higher prevalence of cognitive impairment in the PREVENT sample relative to the VLS. Digit Symbol is highly sensitive to age and the presence of cognitive dysfunction, and it is possible that the increased representation of these characteristics in PREVENT contributed to the sensitivity of RTI T-Scores to performance on Digit Symbol in this sample.

Also consistent with Study 2, a positive linear association between RTI T-Scores and cognitive status was observed in the PREVENT sample. In addition, three of the four examined RTI T-Scores were found to be more important predictors of cognitive status than mean RT. This finding is consistent with the literature documenting the sensitivity of RTI to cognitive impairment and dementia in older adults. The current analyses extend prior research by replicating findings using norm-references measures in a clinical sample. Examination of ROC curves further suggested that RTI T-Scores for the raw ISD, the residual ISD and the MSSD share a small but potentially clinically meaningful association with cognitive status. Consistent with Study 2, continuous RTI T-Scores yielded acceptable sensitivity but specificity was poor for detecting cognitive impairment.

Study 2 found the association between RTI and cognitive status to conform to distribution-based assumptions of cognitive assessment such that the prevalence of impairment was higher in individuals with high RTI T-Scores and lower in examinees with low T-Scores relative to examinees with T-Scores in the average range. This finding was only partly replicated in PREVENT. Consistent with Study 2, the prevalence of cognitive impairment was much lower in participants with RTI T-Scores in the low range relative to examinees with T-Scores in the average range. However, the prevalence of impairment was very similar across participants with average vs. high RTI T-Scores and so there was no significant positive predictive value of a high RTI T-Score in PREVENT data. Indeed, only a very small number (n = 8) of PREVENT examinees yielded RTI T-Scores that were classified in the high range in relation to normative data from the VLS. There is a well-documented linear association between RTI and cognitive impairment, but to date no studies of clinical samples have reported effects of low RTI scores for lowering the odds of impairment. The results of the current study suggest that in clinical samples

the association between RTI and cognitive status may be driven more by the relative protective effect of low RTI than by increased risk of impairment associated with high RTI. However, the small and heterogeneous clinical sample of PREVENT limits the strength of conclusions that can be drawn on the basis of the current findings.

Consistent with the results of Study 2, RTI T-Scores in PREVENT did not predict performance on most tests of physical functioning. A notable exception was the significant univariate association between RTI T-Scores and variability in gait under free and dual-task conditions. Forward regression indicated that the association between gait variability and RTI reflects their shared association with mean RT. However, RTI T-Scores did trend toward independent significance in these analyses, and it is likely that a larger sample size would have elicited statistically significant independent effects of RTI. There is a growing body of literature linking gait speed and variability to cognitive integrity in older adults. Although future research in this area is needed, the results of this study suggest that speed and variability in RT and gait may reflect a common central process that is associated with central nervous system integrity.

The clinical sample in PREVENT remains small (n=26), and heterogeneous due to collapsing of individuals with MCI and AD together for the purpose of maximizing sample size. This is an important limitation because prior research has suggested that the classification of normal vs. impaired examinees differs as a function of impairment severity. For individuals with AD, even simple RT tasks yield RTI scores that distinguish them from healthy controls. In contrast, milder forms of impairment are only distinguishable from healthy aging using RTI scores obtained from executively demanding RT tasks. As the clinical sample of PREVENT increases it will be possible to adopt a more refined approach to classification of cognitive status.

Analyses in this study were carried out using four operationalizations of RTI. Multiple operationalizations of RTI were examined because evidence in the literature is currently lacking regarding the optimal approach to operationalizing RTI for the purpose of detecting cognitive impairment. The results of the current study suggested that the four examined RTI scores yielded generally comparable findings, but subtle differences were observed between measures. For example, linear associations with cognitive status were significant for T-Scores associated with the raw ISD, the residual ISD and the MSSD when examined along with mean RT in forward regression models. In contrast, the CoV did not emerge as a significant independent predictor of cognitive status when included along with mean RT in a forward regression model. This implies that, relative to the other operationalizations of RTI, the association between CoV and cognitive status is more strongly influenced by individual differences in mean RT. In addition, although the raw ISD, residual ISD and MSSD all predicted cognitive status, the residual ISD was found to account for slightly more variance in cognitive status than either the raw ISD or MSSD. This difference, however, was small (e.g., residual ISD T-Score R-Square = 0.19, MSSD T-Score R-Square = 0.15, raw ISD T-Score R-Square = 0.15). Examination of ROC curves suggests even greater agreement among RTI T-Scores for detecting cognitive impairment (e.g., residual ISD T-Score ROC = 0.74, MSSD T-Score ROC = 0.73, raw ISD T-Score ROC = 0.72).

In conclusion, the current study replicated the association between norm-referenced RTI T-Scores and cognitive status in participants of PREVENT. Although replication is warranted in a larger sample, preliminary results suggest that low RTI T-Scores may make a significant independent contribution to the ruling out of cognitive impairment in older adults. Supporting the independent value of measures of RTI relative to other properties of RT data, linear relationships between RTI T-Scores and cognitive outcomes were stronger than those observed for mean RT.

General Discussion

RTI in relation to physical and cognitive function

The results of these studies were generally consistent in the associations that were observed between RTI and physical and cognitive outcomes. The general conclusion across all examined operationalizations of RTI in both samples was that RTI did not yield a significant independent association with physical functioning over and above mean RT. In both VLS and PREVENT data, few significant associations were observed between RTI and performance on tests of physical functioning. Only stride-time variability under free and dual-task conditions yielded an association with RTI in PREVENT, and this association did not retain significance in subsequent forward regression analysis.

Across raw and standardized RTI scores in the VLS and PREVENT, a negative association with cognitive performance was observed such that individuals with higher RTI tended to perform more poorly than individuals with lower RTI. These associations were particularly strong for tests of memory and verbal ability. A discrepancy was observed in results from VLS relative to PREVENT regarding the association between RTI and tests of processing speed. In PREVENT, RTI scores significantly predicted performance visual-motor tests of processing speed. In fact, the Digit Symbol test shared a stronger association with RTI than any other cognitive test in the PREVENT battery. In contrast, in the VLS the Digit Symbol test was the only measure in the VLS cognitive test battery to yield no significant independent association with RTI. As discussed in Study 3, this discrepancy may reflect the differences between studies in overall age of participants, and in the prevalence of cognitive impairment.

Only Study 1 examined the association between RTI and longitudinal change in physical and cognitive function. Contrary to prior research, no association was observed between baseline

SRT or Lexical Decision RTI scores and longitudinal change in cognitive function across three waves of measurement. Similarly, only small and inconsistent associations were observed between SRT RTI scores and longitudinal change in physical functioning. Due to the systematic sequence of analyses in this dissertation, longitudinal change in physical and cognitive function was not examined as a study outcome in subsequent analyses. As discussed in Study 1, prior research examining longitudinal relationships between RTI and cognition has predominantly reported significant correlated change such that longitudinal increases in RTI are associated with longitudinal declines in cognitive performance. The one study to report an association between baseline RTI and longitudinal cognitive change operationalized RTI using a composite score reflecting RTI across a wide range of simple, choice and recognition RT tasks (Macdonald, Hultsch, & Dixon, 2003). It is thus possible that the RT tasks examined in this dissertation were not complex enough to yield prospective associations with cognitive change. This hypothesis is supported by evidence that RT tasks of higher complexity are needed to differentiate subtle impairment from normal performance.

Criterion validity of RTI T-Scores for detecting cognitive impairment

The main objective of this dissertation was to determine whether RTI scores can be norm-referenced and applied clinically to detect cognitive impairment. To address this goal, data from cognitively intact Sample 2 participants of the VLS were stratified by age and education and used as a normative reference group against which PREVENT and Sample 3 VLS data were standardized. Normative data were stratified because RTI scores in the standardization sample were observed to differ systematically as a function of age and education. Were demographic stratification not implemented, standardized RTI scores would have been biased against older participants and those with limited education.

RTI T-Scores from Sample 3 of the VLS were found to yield associations with cognitive performance and cognitive status that were of a similar magnitude as those observed from raw RTI scores in the same dataset. This finding was replicated in PREVENT, where Raw ISD and Residual ISD T-scores yielded larger ROC curve values and accounted for more variance in cognitive status than Mean RT. Although associations between RTI and performance on individual neuropsychological tests attenuated following adjustment for Mean RT, associations between continuous RTI T-Scores and the main study outcomes for cognitive status were found to be independent of Mean RT. These results replicate and extend the results of many other studies that have documented the sensitivity of measures of RTI to cognitive status.

In addition to examining RTI T-Scores as linear variables, this dissertation also examined categorical RTI T-Score variables classifying examinees as possessing RTI scores in the "low", "average" or "high" range relative to others of their age and educational background. This approach is in keeping with clinical assessment procedures, where a distributional approach is used to classify performance as normal vs. impaired. As is discussed in Study 2 and Study 3, a discrepancy was observed between results from VLS and PREVENT such that no positive predictive value of "high" RTI scores was observed in PREVENT. It is possible that the VLS normative sample was not sufficiently representative of PREVENT examinees, and the obtained T-Scores may thus have been biased in a way that obfuscated the association between elevated RTI T-Scores and cognitive impairment. However, this is unlikely because the distributional characteristics of the RTI T-Scores were similar in the two samples. It is also possible that the small sample of PREVENT contributed to the discrepancy of findings. As PREVENT data collection continues, the direction of these findings may adjust to conform more closely with VLS data. Finally, the operationalizations of cognitive status in VLS and PREVENT differed,

and PREVENT included a heterogeneous sample of participants with a wide range of impairment severity. It is thus possible that the different characteristics of the clinical samples between groups contributed to the discrepant findings. The results of this dissertation extend prior research by demonstrating that norm-referenced RTI scores have the potential to make a small but meaningful independent contribution to the detection of cognitive impairment.

Optimal computation of RTI for clinical use

The inconsistent operationalization of RTI in the literature necessitated an examination of multiple approaches to the computation of RTI. The wide range of RTI computations that are reported in the literature largely reflect attempts to address the overlap between RTI and mean reaction time (RT) in relation to CNS dysfunction. General slowing is itself sensitive to CNS dysfunction of many etiologies, and because higher mean RT values naturally yield higher standard deviations it is challenging to demonstrate associations between RTI and CNS integrity that can be proven independent of mean RT. As reviewed by Dykiert and colleagues (2012), the direction of the causal association between mean RT and RTI is unclear: increases in RTI may purely reflect increases in mean RT, increases in mean RT may be driven by neural processes that exert direct influences on RTI, or there may be no causal association between mean RT and RTI, with these two properties representing independent manifestations of CNS dysfunction. The hypothesis that RTI is driven purely by increases in mean RT is not supported by the current study, where associations between RTI and cognitive performance were observed in models adjusted for mean RT, age, sex and educational attainment. In support of the behavioural findings reported here and elsewhere, research using fMRI and PET has demonstrated unique neural correlates of RTI and Mean RT in healthy older adults.

This dissertation found the raw ISD, the residual ISD and the MSSD to yield highly similar associations with cognitive function and cognitive status. The CoV also correlated strongly with these other operationalizations and yielded many associations with study outcomes that were of a similar direction and magnitude as the MSSD and the raw and residual ISD. The general pattern of findings indicated that the residual ISD shares slightly stronger associations with cognitive outcomes relative to the other examined RTI scores. However, for the purposes of this dissertation, which are to determine the clinical utility of RTI scores, validity issues must be balanced with issues of feasibility. The residual ISD is more computationally intensive than the MSSD or certainly the raw ISD, and this dissertation did not find any clinically meaningful difference in the strength of the association of these three scores with cognitive outcomes. Thus, the raw ISD and the MSSD may be preferable operationalizations for clinical use. Of these two, MSSD is more computationally demanding but could easily be computed using a basic script. The MSSD is also conceptually preferable to the raw ISD because it adjusts for gradual shifts in latency that could be attributable to fatigue or practice. However, this study found slightly stronger and more consistent associations between the raw ISD and study outcomes relative to the MSSD. Further study is warranted in order to determine whether there are any circumstances (e.g., very old and more fatigable populations) in which the MSSD would meaningfully outperform the raw ISD in detecting cognitive dysfunction.

The results of this dissertation provide support for the clinical utility of low RTI scores for establishing older adults as cognitively intact and ruling out the presence of impairment. Likelihood ratios suggest that consideration of RTI scores can make a small put potentially meaningful change in the probability of accurately detecting cognitive impairment. The magnitude of this effect does not suggest that RTI scores should serve as a standalone tool for

detecting or ruling out the presence of cognitive impairment. However, it suggests that RTI scores, used in conjunction with other sources of information, are likely to contribute meaningfully to clinical decision making. In recognition of the association between RTI and Mean RT, such that examinees with fast mean RT values would be expected to have smaller RTI scores than those with slower mean RT values, it would be prudent to interpret RTI scores in the context of an examinee's mean RT.

Study limitations and directions for future research

There are several potential limitations to the results reported in this dissertation. Main limitations include the operationalization of cognitive status in VLS and PREVENT, the small clinical sample size of PREVENT, the limited number of RT trials available for computation of RTI scores, increased possibility of Type I error due to the high number of statistical comparisons that are reported, and the case-control design of PREVENT. These limitations present opportunities for future research to replicate and extend the findings of this dissertation.

The VLS and PREVENT were both designed to address research questions that were not directly related to the objectives of the current dissertation. The VLS, as a study of healthy aging, was not designed for the purpose of studying cognitive impairment or dementia, or for the validation of tools for use with clinical populations. Participants of the VLS are a healthier and higher-functioning population than would be expected of a clinical sample. In addition, longitudinal cognitive change in the VLS is likely to reflect declines associated with normal, as opposed to pathological, cognitive aging. Operationalizations of cognitive impairment that were examined in the VLS were distribution-based, and did not necessary include a subjective report of cognitive decline or even a decline from a previous level of functioning. Thus, the operationalizations of cognitive impairment that were examined in Study 1 and Study 2 were

experimental in nature and cannot be assumed to correspond with a clinical diagnosis of cognitive impairment. For the purpose of the current dissertation – to determine the clinical utility of RTI – the experimental nature of the operationalization of cognitive status is a limitation because results from the VLS cannot be assumed to generalize to a clinical population.

This dissertation was structured such that the conceptual limitations of the VLS sample were intended to be addressed by replication of analyses in PREVENT. However, sample size limitations of PREVENT represented an added limitation that may compromise conclusions that can be drawn from the results of this dissertation. PREVENT currently includes 26 participants who met study criteria for cognitive impairment. Of these participants, 7 met criteria for probable AD and the remaining 19 participants fell within a milder spectrum of cognitive. While the heterogeneity of impairment in PREVENT is representative of real-life clinical practice, prior research has shown that the association between RTI and cognitive impairment may differ as a function of impairment severity, and thus the results of Study 3 may have differed had more homogenous samples of impaired participants been available.

The majority of research examining RTI in relation to cognitive status has been obtained from RT tasks involving fewer than 100 trials. The number of trials examined in the current dissertation (60 trials for the Lexical Decision task, 50 trials for the SRT) are thus consistent with prior research. However, operationalizations of RTI that are obtained from distribution-based parameters (e.g., Ratcliff, Ex-Gaussian parameters) require more trials than were available in either the VLS or PREVENT and thus the potential clinical utility of these operationalizations could not be examined in this dissertation. Prior research has suggested that associations between Ex-Gaussian distributions and RTI operationalizations that were examined in this study (e.g., ISD, CoV) are strong, and associations with cognitive outcomes are comparable. It is thus

unlikely that the unexamined parameters would have yielded meaningfully stronger associations with the cognitive outcomes in this study.

Studies 1, 2 and 3 were carried out in a sequential manner with an emphasis on replication of findings in independent datasets. However, particularly for Study 1, multiple operationalizations of RTI obtained from two different RT measures were examined in relation to many physical and cognitive outcomes. To reduce the likelihood of Type I error, significant results were interpreted with attention to the consistency of the pattern of significant vs. nonsignificant findings. In addition, a more conservative *p*-value of <=0.01 was used to correct for multiple comparisons.

Classification accuracy statistics assume that the prevalence of a given disorder in a research sample is equivalent to the prevalence of the disorder in the general population. The proportion of VLS participants who were classified as impaired (11%) was roughly in line with the base rate of cognitive impairment in the general population. However, PREVENT, as a study that follows a case-control design, has a clinical sample that accounts for 34% of the full study sample. It is thus possible that estimates obtained from PREVENT data may be inflated by the high prevalence of cognitive impairment in this sample.

This dissertation describes the first systematic attempt to develop norm-referenced measures of RTI and evaluate their feasibility and validity of for clinical use. Preliminary evidence suggests that RTI scores obtained from choice RT tasks can be standardized and applied clinically to aid in detection of cognitive impairment. Low RTI T-Scores were shown across two independent samples to be associated with a clinically meaningful reduction in the odds of cognitive impairment independent of mean RT. Further research is needed in order to clarify the utility of high RTI scores for positive prediction of cognitive impairment.

References

Akobeng, A. K. (2007), Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatrica, 96: 338–341.

American Psychological Association (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Anstey, K. J., Dear, K., Christensen, H., & Jorm, A. F. (2005). Biomarkers, health, lifestyle, and demographic variables as correlates of reaction time performance in early, middle, and late adulthood. *Q.J.Exp.Psychol.A, 58*(1), 5-21.

Anstey, K. J., Mack, H. A., Christensen, H., Li, S. C., Reglade-Meslin, C., Maller, J., . . . Sachdev, P. (2007). Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia, 45*(8), 1911-1920.

Bielak, A. A., Hultsch, D. F., Strauss, E., Macdonald, S. W., & Hunter, M. A. (2010). Intraindividual variability is related to cognitive change in older adults: Evidence for within-person coupling. *Psychol.Aging, 25*(3), 575-586.

Bunce, D., Anstey, K. J., Christensen, H., Dear, K., Wen, W., & Sachdev, P. (2007). White matter hyperintensities and within-person variability in community-dwelling adults aged 60-64 years. *Neuropsychologia, 45*(9), 2009-2015.

Cherbuin, N., Sachdev, P., & Anstey, K. J. (2010). Neuropsychological predictors of transition from healthy cognitive aging to mild cognitive impairment: The PATH through life study. *Am.J.Geriatr.Psychiatry, 18*(8), 723-733.

Christensen, H., Dear, K. B., Anstey, K. J., Parslow, R. A., Sachdev, P., & Jorm, A. F. (2005). Within-occasion intraindividual variability and preclinical diagnostic status: Is intraindividual variability an indicator of mild cognitive impairment? *Neuropsychology, 19*(3), 309-317.

de Frias, C. M., Dixon, R. A., Fisher, N., & Camicioli, R. (2007). Intraindividual variability in neurocognitive speed: A comparison of parkinson's disease and normal older adults. *Neuropsychologia, 45*(11), 2499-2507.

Dixon, R. A., Garrett, D. D., Lentz, T. L., Macdonald, S. W., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology, 21*(3), 381-399.

Duchek, J. M., Balota, D. A., Tse, C. S., Holtzman, D. M., Fagan, A. M., & Goate, A. M. (2009). The utility of intraindividual variability in selective attention tasks as an early marker for alzheimer's disease. *Neuropsychology, 23*(6), 746-758. doi: 10.1037/a0016583; 10.1037/a0016583

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.

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012b). Sex differences in intra-individual variability in simple and choice reaction time: Systematic review and meta-analysis. *PLoS One, 7*(10), e45759.

EClipSE Collaborative Members (2010). Education, the brain and dementia*:* Neuroprotection or compensation*? Brain*, *133*(8), 2210-2216.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests.* Princeton, NJ: Educational Testing Service.

Fjell, A. M., Westlye, L. T., Amlien, I. K., & Walhovd, K. B. (2011). Reduced white matter integrity is related to cognitive instability. *J.Neurosci., 31*(49), 18060-18072.

Gorus, E., De, R. R., Lambert, M., Lemper, J. C., & Mets, T. (2008). Reaction times and performance variability in normal aging, mild cognitive impairment, and alzheimer's disease. *J.Geriatr.Psychiatry Neurol., 21*(3), 204-218.

Greenhalgh, T. (2010). *How to Read a Paper: The Basics of Evidence-Based Medicine.* Chichester: Wiley-Blackwell (4th Ed.).

Hanley, J. A., McNeil ,B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology,* 143, 29-36.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin, 109*, 340-347.

Hultsch, D. F., Macdonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *J.Gerontol.B Psychol.Sci.Soc.Sci., 57*(2), 101-115.

Hultsch, D. F., Macdonald, S. W., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14*(4), 588-598.

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.

Jaeschke, R., Guyatt, G. H., Sackett, D. L. (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA*, 271(9), 703-707.

Lovden, 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.

Lovden, M., Schmiedek, F., Kennedy, K. M., Rodrigue, K. M., Lindenberger, U., & Raz, N. (2013). Does variability in cognitive performance correlate with frontal brain volume? *NeuroImage, 64*, 209-215. doi: 10.1016/j.neuroimage.2012.09.039; 10.1016/j.neuroimage.2012.09.039

Macdonald, S. W., Cervenka, S., Farde, L., Nyberg, L., & Backman, L. (2009). Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. *Neuropsychologia, 47*(11), 2299-2304.

Macdonald, S. W., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychol.Aging, 18*(3), 510-523.

Macdonald, S. W., Hultsch, D. F., & Dixon, R. A. (2008). Predicting impending death: Inconsistency in speed is a selective and early marker. *Psychol.Aging, 23*(3), 595-607.

MacDonald, S. W., Nyberg, L., & Backman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences, 29*(8), 474-480. doi: 10.1016/j.tins.2006.06.011

Martin, K. L., Blizzard, L., Wood, A. G., Srikanth, V., Thompson, R., Sanders, L. M., Callisaya, M. L. (2013). Cognitive function, gait, and gait variability in older people: a population-based study. *J Gerontol A Biol Sci Med Sci, 68*(6), 726-732.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Reviews, 16*, 798-817.

Moy, G., Millet, P., Haller, S., Baudois, S., de, B. F., Weber, K., . . . Delaloye, C. (2011). Magnetic resonance imaging determinants of intraindividual variability in the elderly: Combined analysis of grey and white matter. *Neuroscience, 186*, 88-93.

Murtha, S., Cismaru, R., Waechter, R., & Chertkow, H. (2002). Increased variability accompanies frontal lobe damage in dementia. *J.Int.Neuropsychol.Soc., 8*(3), 360-372.

Nunnally, J. & Bernstein, I. (1994) *Psychometric Theory.* New York: McGraw Hill, 3rd ed.

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychol.Aging, 23*(4), 900-916.

Salthouse, T.A. (2012). Psychometric properties of within-person across-session variability in accuracy of cognitive performance. *Assessment*, 19, 494-502.

Santhanam, P., Simon, J.R., Seaman, K.L., Howard, D.V., & Howard, J.H. Jr. (2013, May). Brain and behavioral intra-individual variability in aging during an implicit learning task. *Presented at the annual Association for Psychological Science Convention.* Washington, DC.

Schmiedek, F., Lövdén, M., Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology & Aging, 24*(4)841-57.

Spirduso, W. W., & Clifford, P. (1978). Replication of age and physical activity effects on reaction and movement time. *J.Gerontol., 33*(1), 26-30.

Strauss, E., Bielak, A. A., Bunce, D., Hunter, M. A., & Hultsch, D. F. (2007). Within-person variability in response speed as an indicator of cognitive impairment in older adults. *Neuropsychol.Dev.Cogn B Aging Neuropsychol.Cogn, 14*(6), 608-630.

Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain: A Journal of Neurology, 126*(Pt 11), 2363-2380. doi: 10.1093/brain/awg237

Tales, A., Leonards, U., Bompas, A., Snowden, R. J., Philips, M., Porter, G., . . . Bayer, A. (2012). Intra-individual reaction time variability in amnestic mild cognitive impairment: A precursor to dementia? *J.Alzheimers Dis., 32*(2), 457-466.

Thurstone, T. G. (1962). *Primary mental abilities: Grades 9–12, 1962 revision.* Chicago: Science Research Associates.

Vandermorris, S., Hultsch, D. F., Hunter, M. A., MacDonald, S. W., Strauss, E. (2011). Including persistency of impairment in Mild Cognitive Impairment classification enhances prediction of 5-year decline. *Archives of Clinical Neuropsychology*, *26(1)*, 26-37.

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

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain Cogn, 49*(3), 402-419.

Whitehead, B. P., Dixon, R. A., Hultsch, D. F., & Macdonald, S. W. (2011). Are neurocognitive speed and inconsistency similarly affected in type 2 diabetes? *J.Clin.Exp.Neuropsychol., 33*(6), 647-657.