Journal of School Psychology 50 (2012) 841–863

Contents lists available at SciVerse ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

# A dynamic developmental link between verbal comprehension-knowledge (Gc) and reading comprehension: Verbal comprehension-knowledge drives positive change in reading comprehension☆

Matthew R. Reynolds*, Joshua J. Turek

*The University of Kansas, United States*

ARTICLE INFO

*Article history:*
Received 15 November 2011
Received in revised form 27 June 2012
Accepted 5 July 2012

*Keywords:*
Bivariate dual change score models
Reading comprehension
Crystallized intelligence
Verbal-comprehension knowledge
CHC theory
Intelligence

ABSTRACT

Intelligence and general academic achievement have a well-established relation, but the interrelated development of the two constructs over time is less well-known. In this study, the dynamic developmental relation between verbal comprehension-knowledge (Gc) and reading comprehension was examined by applying bivariate dual change score models (McArdle, 2009) to longitudinal data collected from children aged 9 through 15 who were part of the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD). A unidirectional dynamic link was found in which higher levels of prior Gc led to increased positive change in reading comprehension scores. This unidirectional link was not altered by including intelligence measured at 24-months, SES, sex, basic reading, and reading volume as time-invariant covariates. Gc is a leading indicator of reading comprehension and should be considered when developing and monitoring long-term reading comprehension interventions for children.



☆ This study was conducted by the NICHD Early Child Care Research Network supported by NICHD through a cooperative agreement that calls for scientific collaboration between the grantees and the NICHD staff. We appreciatively acknowledge the data provided by this study and all of those involved in the data collection process.

* Corresponding author at: University of Kansas, Psychology and Research in Education, Joseph R. Pearson Hall, 1122 West Campus Rd, Lawrence, KS 66045-3101, United States. Tel.: +1 785 864 9712.

*E-mail address:* mreynolds@ku.edu (M.R. Reynolds).

ACTION EDITOR: Renee Hawkins.



doi:10.1016/j.jsp.2012.07.002

## 1. Introduction

The relation between intelligence and academic achievement is well-established. Intelligence scores are correlated with academic achievement when the two are measured at the same time (Gustafsson & Balke, 1993; Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012; Rohde & Thompson, 2007). Intelligence scores also predict future academic achievement (Deary, Strand, Smith, & Fernandes, 2007). Although it is evident that intelligence and academic achievement are related, an understanding of the dynamic, developmental relations among intelligence constructs and specific academic skills is in its nascency.

### *1.1. Stability and change in intelligence and academic achievement*

The rank ordering of IQ scores is remarkably stable over time (Deary, Whalley, & Starr, 2009). From a developmental point of view; however, intelligence is not "fixed" because although the rank ordering of scores is generally stable (relative continuity), within-individual changes (absolute continuity) also occur (Cattell, 1963; McArdle, Hamagami, Meredith, & Bradway, 2000). For example, a 5-year-old child with an IQ score of 100 does not have the same level of intelligence as a 20-year-old adult with an IQ score of 100, or in general, a raw score on an IQ test for someone who took the test when they were 5 would be less than the raw score on the same IQ test if it were taken again when they were 20. Childhood and adolescence, in particular, is a time of rapid cortex development, and children with higher general intelligence demonstrate more dynamic cortical changes (Shaw et al., 2006). During this time, domain general intelligence scores, including those related to crystallized intelligence and fluid intelligence, also demonstrate concomitant rapid growth (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002) such that changes in the absolute level of scores are likely reflecting changes in the cortex.

The rank ordering of academic achievement scores is also generally stable over time (Verhoeven & Van Leeuwe, 2008). Childhood and adolescence is a time when academic achievement skills are acquired rapidly. Math, reading, and writing skills show rapid growth during elementary school (McArdle et al., 2002), which also coincides with children's rapid increases in intelligence scores, cortical maturation, and exposure to academic content via formal education.

Although within-construct intelligence and academic achievement stability correlations are strong, the development of intelligence and the acquisition of academic achievement skills are not believed to occur in a vacuum. They are thought to be dynamically interrelated over time (Cattell, 1987). It is thus important to investigate stability and change, within and across constructs, and within and across individuals, when studying the development of intelligence and the acquisition of academic achievement skills. Except for a few studies (e.g., Ferrer, Hamagami, & McArdle, 2004; Ferrer & McArdle, 2004), there has been limited research investigating the dynamic, developmental relation between intelligence and academic achievement over time. The purpose of this study is to model the developmental dynamics of children's verbal comprehension-knowledge (Gc) and reading comprehension in grades 3 through 9.

### *1.2. Verbal comprehension-knowledge (Gc)*

The term "Gc" is often associated with Cattell's (1987) investment hypothesis, which was an aspect of his fluid ($g_f$) and crystallized ($g_c$) theory of intelligence (Cattell, 1963, 1987; also see Hebb, 1942; Horn & Cattell, 1966). Cattell hypothesized a biologically and physiologically determined general relation-perceiving capacity. This general capacity, $g_f$, is invested into the acquisition of new skills and knowledge across domains. As these new skills are acquired via exposure and practice, they become organized and crystallized in the cortex. These crystallized abilities are then also involved in the future acquisition of academic achievement (Cattell, 1963).

Conceptualization and measurement of Gc, however, has not always been explicitly tied to the investment hypothesis.[1] In general, Gc, as measured by popular, individually administered measures of intelligence, appears to represent verbal comprehension-knowledge ability (Kan, Kievit, Dolan, & van der Maas, 2011; Schneider & McGrew, 2012). Specifically, most vocabulary and verbal knowledge is acquired when verbal comprehension and reasoning abilities are used to educe correlates and generalize stored knowledge to newly encountered verbal information (cf. Carroll, 1993; Flanagan, Ortiz, Alfonso, &

[1] Research has not necessarily supported several aspects of this hypothesis (Horn, 1991).

Mascolo, 2006; Guttman, 1965; Horn & McArdle, 2007; Jensen, 1980). Cross-sectional research has consistently demonstrated a first-order Gc factor that is reliably different from a second-order general factor when data are analyzed in hierarchical models of intelligence (Carroll, 1993).[2] Controlling for this general factor, Gc is also independent of several other broad cognitive factors, such as Visual Processing and Short-Term Memory. Cattell–Horn–Carroll theory (CHC), a working model of intelligence based on hundreds of factor-analytic studies, and an amalgamation of Carroll's three-stratum theory (Carroll, 1993) and Gf–Gc theory, includes a Gc factor, which is described as verbal comprehension-knowledge (Schneider & McGrew, 2012). Thus, Gc is hypothesized as a common cause of correlations among scores from measures of vocabulary knowledge, verbal proficiency, and general verbal knowledge, or quite simply, it produces variation in scores from measures of those skills. Gc shows rapid absolute growth in early childhood and adolescence, with growth up until the mid-30s, stability well into late adulthood, and average decline not observed until the early-70s (McArdle et al., 2002).

Gc, especially during the years when children and adolescents are encountering formal academic instruction, is conceptually distinct from academic skills, although the two are often used interchangeably (Cattell, 1963; Keith & Reynolds, 2010). For example, CHC theory includes academic-like factors such as reading and writing (Grw) and quantitative knowledge (Gq), which are correlated with, but reliably distinct from a Gc factor. Moreover, measurement of Gc in intelligence batteries also differs from the measurement of subject-matter-specific academic achievement skills that are taught formally in school. Academic skills (e.g., math, reading, and writing) are typically measured in a separate academic achievement battery (e.g., the Woodcock–Johnson III Tests of Achievement; Woodcock, McGrew, & Mather, 2001, 2007, and the Kaufman Tests of Educational Achievement — Second Edition; Kaufman & Kaufman, 2004).

In general, ability constructs, such as Gc, are believed to cut across content domains, whereas academic achievement is subject matter specific (Messick, 1984). Hence, Gc is thought of as a causal variable that underlies the acquisition of vocabulary and verbal knowledge across domains. Thus, Gc is a reflective latent variable, such that an increase in Gc should produce increases in the scores on all of its indicators (Borsboom, Mellenbergh, & Van Heerden, 2003). Academic achievement constructs, however, are likely more appropriately conceptualized as emergent or formative variables, which simply summarize what has been learned in formal schooling. An increase in general achievement would not produce increases in all achievement indicators, rather an increase in a specific academic achievement skill would produce an increase in general achievement, while the other indicators would be unaffected. In this study, we ascribe to the view that Gc is a reflective latent verbal comprehension-knowledge and reasoning ability (Kan et al., 2011) that is related to, but independent from, academic achievement.

### *1.3. Reading comprehension, Gc, and vocabulary*

Reading comprehension is generally defined as a skill to derive meaning from text. In the parlance of the CHC theory, reading comprehension represents an indicator of a reflective latent variable Grw (Woodcock, 2002).[3] Reading comprehension skills are often estimated from scores taken from measures requiring individuals to read a passage aloud or silently and then answer questions via direct recall or multiple-choice. It is also measured by cloze tasks that require the reader to use context cues from a passage to identify words that are missing from the text. Reading comprehension scores are relatively stable over time (Betjemann et al., 2008). On average, reading comprehension skills increase rapidly during the years of required formal schooling, peak in the mid-20s, and decline slowly beginning around the age of 50 (McArdle et al., 2002; McGrew, Schrank, & Woodcock, 2007).

From a theoretical standpoint, Gc should affect reading comprehension, although the relation is often considered reciprocal (Stanovich, 1986). In addition, more specific aspects of Gc, vocabulary and prior knowledge, are believed to underlie reading comprehension because they allow the reader to connect what they know with information presented in a text (Cain, Oakhill, & Bryant, 2004). From an

[2] Not all researchers ascribe to the existence of a general factor; such discussion is beyond the purpose of this research (cf. Carroll, 1993; Horn & McArdle, 2007).

[3] Grw may also be viewed as a *composite variable*, which summarizes several different skills such as word reading, writing, and spelling, and not as a reflective latent variable from which the causal paths flow into to the subtests.

empirical standpoint, several lines of research, including cross-sectional and longitudinal studies and studies of individuals with reading comprehension deficits, support a relation between Gc and reading comprehension.

A recent synthesis of empirical research on the influence of CHC abilities and academic achievement concluded that Gc, as well as its specific aspects, including language development, general verbal information, and vocabulary knowledge, are related to reading comprehension for children and adolescents aged 8 to 19 (McGrew & Wendling, 2010). The relation is either consistently strong, or the influence of Gc increases during the years of formal schooling (cf. Evans, Floyd, McGrew, & Leforgee, 2002; Perfetti, Liu, & Tan, 2005; Rupley, Willson, & Nichols, 1998; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009). The influence of Gc on reading comprehension has also been demonstrated in studies in which researchers utilized structural equation models, which in turn allowed for different aspects of intelligence to be isolated when studying the effects of intelligence on reading comprehension. Benson (2008) demonstrated that a latent Gc ability had a direct effect on reading comprehension, which was very strong in grades 4 through 12. Keith (1999) found a strong and direct latent Gc effect on reading comprehension from grades 1 through 12, with this relation being for the most part similar across various ethnic groups.

Additional evidence for a Gc and reading comprehension link comes from research on individuals who struggle with reading comprehension. A group of readers referred to as “poor comprehenders,” who include about 5% to 10% of school-aged children (Nation & Snowling, 1997), show deficits in reading comprehension despite intact word recognition skills (Catts, Adolf, & Ellis Weismer, 2006; Nation, Clarke, & Snowling, 2002). This group of readers has often been characterized by language deficits, including difficulty with semantic processing. Moreover, cognitive profiles of children with reading comprehension deficits indicate concurrent deficits in Gc (Floyd, Bergeron, & Alfonso, 2006). Clearly, Gc and reading comprehension are closely linked.

Specific aspects of Gc (i.e., CHC narrow abilities) also show relations with reading comprehension. Vocabulary knowledge, for example, perhaps the best indicator of Gc, has a strong relation with reading comprehension. This correlation tends to increase throughout elementary school. Three hypotheses have been generated to possibly explain this correlation (Anderson & Freebody, 1981). The first, the *instrumentalist hypothesis*, is that there is a causal relation between vocabulary knowledge and reading comprehension, such that increasing vocabulary will increase reading comprehension. The second, the *knowledge hypothesis*, suggests that stored knowledge is related to higher vocabulary and reading comprehension, such that increases in stored general knowledge, will increase vocabulary and reading comprehension. The third, the *aptitude hypothesis*, indicates that there is a common cause (e.g., general intelligence) that produces the correlation between vocabulary and reading comprehension scores. It is plausible that all three of these are correct to some extent when considered longitudinally. For example, knowledge may mediate the effects of general intelligence on vocabulary knowledge and reading comprehension. Rarely however have such, or other dynamic-type, relations been mapped formerly on a longitudinal statistical framework.

### *1.4. Dynamic relations between IQ scores, Gc, and reading*

Research on the relation between IQ scores, Gc, and reading constructs has focused mostly on static relations. The developmental relation between aspects of intelligence and reading, however, has been examined longitudinally within the context of Cattell's investment hypothesis. These studies have been enlightening because the statistical models have incorporated aspects of development that unfold in a time-ordered sequence, while also capturing how one developmental process is dynamically interwoven with other developmental processes (e.g., Ferrer & McArdle, 2004; Ferrer, Shaywitz, Holahan, Marchione, & Shaywitz, 2010; McArdle et al., 2000). Using data from the Connecticut Longitudinal Study (Shaywitz, Shaywitz, Fletcher, & Escobar, 1990), Ferrer et al. (2007) applied bivariate latent dual change score models to study the dynamic relations between reading and intelligence in childhood through late adolescence. Verbal IQ, derived from the Wechsler Intelligence Scales for Children — Revised (WISC-R; Wechsler, 1974) scales, was considered an estimate of Gc in these models. They found, among other things, that Verbal IQ and reading comprehension maintained a statistically significant and mutually beneficial dynamic relation over time, so that higher levels of one variable measured at a previous time point resulted in more positive changes in the other variable measured at a later time point.

In another study, with a sample ($N=232$) of participants from the Connecticut Longitudinal Study (Shaywitz et al., 1990), Ferrer et al. (2010) studied the dynamic relation between Full Scale IQ (FSIQ) and Reading Cluster scores from the Woodcock–Johnson Psychoeducational Battery (WJ; Woodcock & Johnson, 1977). They compared models across a sample of typical readers, readers who were persistently poor, and readers who later compensated for poor reading fluency (i.e., whose reading comprehension scores were not poor in grades 9 or 10). They found mutually beneficial relations between FSIQ and reading over time in typical readers but not in the other reading groups. In the other two groups, prior FSIQ had only small positive influences on changes in reading over time, but the influence in the compensated reading group was larger than in the poor reading group. In these two groups, prior levels of reading did not influence changes in FSIQ over time. These findings also generalized to models that included individual reading tests rather than the Reading Cluster scores. Last, it was interesting to note that in the compensated reader group, FSIQ scores initially were higher than reading scores, but by adolescence the reading comprehension scores were similar to those of typical readers, despite poorer reading fluency. On the other hand, persistently poor readers had similarly low initial FSIQ and reading scores. The findings thus revealed that initial levels of IQ may play an important role in the development of reading comprehension for children who struggle with reading comprehension initially.

### *1.5. Present study*

#### *1.5.1. Goals of the study*

There is sufficient evidence of a relation between reading comprehension and Gc (McGrew & Wendling, 2010). Less is known about the developmental dynamics between these constructs. We will investigate the dynamic interplay between Gc and reading comprehension in children from grades 3 through 9, with two general goals in mind. First, we will attempt to replicate the seminal work of Ferrer et al. (2007, 2010) with a different sample of participants and measures. We will test the dynamic hypotheses in a large longitudinal sample that is independent from the Connecticut Longitudinal Study, and we will utilize Rasch-based Gc and reading comprehension scores, from a co-normed intelligence and achievement test battery, which are optimal for longitudinal research. To reach this goal, we will apply bivariate dual change score models (Ferrer & McArdle, 2004; McArdle, 2009; McArdle & Hamagami, 2001). These models were considered appropriate because of their consistency with conceptualizations of intelligence and academic achievement development in that there is dual change within a construct, additive (constant) and proportional growth, as well as dynamic changes among the constructs. Bivariate dual change models incorporate aspects of latent curve models and cross-lag regression models, overcoming limitations of these methods used in isolation (McArdle, 2009). For example, latent curve models are restricted in the sense that they do not typically model autoregressive effects; moreover, dynamic hypotheses between constructs cannot be tested using latent curve models (cf. Bollen & Curran, 2006). Cross-lag regression models do not account for changes in means over time (absolute change), and it is well-known that both Gc and reading comprehension grow over time.

Second, we will attempt to expand on the previous research by including a variety of background and mediating variables to estimate their total, indirect, and direct influences on the initial levels and slopes of Gc and reading comprehension. We included early childhood intelligence, measured at 24 months of age, and socio-economic status (SES), indexed by maternal education, as background variables because we believe that a general early-relation-perceiving capacity that develops in infancy and toddlerhood and access to resources that promote learning likely explain variation in 3rd-grade Gc and reading comprehension scores as well as within-person growth in these constructs. In addition, because there may have been some heterogeneity in Gc and reading comprehension scores induced by biological sex, sex was included as a background variable.

In addition to background variables, relative reading volume and basic sight word recognition were included as mediating variables (Cunningham & Stanovich, 2001). Some researchers have indicated that reading volume is a primary driver of vocabulary growth. That is, vocabulary knowledge is not acquired via direct vocabulary instruction; rather, it is acquired indirectly, via reading or exposure to language (Nagy, Herman, & Anderson, 1985). Others have indicated that vocabulary knowledge is mostly developed via reading because verbal interactions contain few words that are outside of one's current vocabulary, especially after 6th grade. Thus, exposure to print will improve Gc (Cunningham & Stanovich, 2001). Similarly, basic word recognition skills are related to reading comprehension and the amount of material read, which in turn should improve vocabulary or general knowledge. Estimates of relative reading volume and basic word reading skills were therefore also included in this study. Specifically, these variables were likely mediators of background variable influences on Gc and reading comprehension.

The current study is of both theoretical and practical import. Understanding the interrelated longitudinal development of these processes is essential for improved and more refined theory. Understanding the interplay and development should help inform psychologists and educators in setting parameters of expectations for child learning. Moreover, understanding reading comprehension within a developmental context may lead to new or more informed interventions.

*1.5.2. Research questions and hypotheses*

The purpose of the study was to describe the development of Gc and reading comprehension from 3rd grade through 9th grade using bivariate dual change score models. We intended to answer two broad questions:

1. The primary research question was whether there was a dynamic interrelated process between Gc and reading comprehension?
   a. If yes, was this process mutually beneficial? Did Gc drive changes in reading comprehension? Or, did reading comprehension drive changes in Gc?
2. What was the role of important background and mediating variables in the development of Gc and reading comprehension?

Because intelligence and academic achievement scores increase during childhood and adolescence, with particularly steep growth observed during the early elementary years, it was expected that both Gc and reading comprehension trajectories would follow an increasing yet decelerating form (McArdle et al., 2002). In the dual change score model, the expected average trajectory was expected to be defined by a positive linear slope factor mean representing a constant increase, with negative autoproportions representing deceleration. Based on prior research, we expected a mutually beneficial relation between Gc and reading comprehension over time (Ferrer et al., 2007).

SES and early childhood intelligence were expected to influence 3rd grade Gc and reading comprehension scores, with higher SES and early childhood intelligence leading to higher scores. SES and early childhood intelligence were also expected to influence internal Gc growth and possibly reading comprehension growth, with higher SES and early childhood intelligence associated with steeper increases in growth. The expected effects of sex were unknown, although differences were not expected to be large.

Basic sight word reading skill was hypothesized to influence initial levels of, and possibly, growth in reading comprehension. Its influence on Gc was unknown. Reading volume was expected to influence initial levels of and growth in reading comprehension and Gc. Moreover, these two variables were expected to mediate some of the total effects from the background variables, so they were modeled accordingly.

## 2. Method

The National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD) was an extensive longitudinal study and has led to a large body of developmental research (NICHD Early Child Care Research Network, 1993). This multisite study was conducted in 10 different locations throughout the United States beginning in 1991. Inclusion criteria requirements were that the mother of the targeted child was age 18 years or older, spoke English, and lived within an hour of a research location in a neighborhood not considered too dangerous for visits by research staff. It was also required that the child was not part of a multiple birth or adopted. A total of 1525 families were eligible, and 1364 completed a home interview when the child was 1 month of age.

The NICHD participant recruitment occurred in the first 11 months of 1991, with 8986 mother–infant dyads originally screened. The original study sample was obtained through use of a conditionally random selection procedure, in an attempt to include at least 10% of each of the following demographic characteristics: single parent households, mothers with less than a high school education, and mothers identifying as ethnic minorities. Enrolled families at each of the 10 sites included approximately 60% of mothers planning to work full time during the next year, 20% part time, and 20% not anticipating to work in the next year. For further details regarding study sampling and site information refer to various manuscripts from the NICHD Early Childhood Care Research Network (ECCRN) or http://www.nichd.nih.gov/research/supported/seccyd.cfm.

### 2.1. Participants

Children in the NICHD SECCYD were assessed at various time points from 1 month to 15 years of age. The sample for the current study included 1079 participants, comprising the active participants in 3rd grade, the first time point of interest. The assessments were for the most part based on grade level and not chronological age. The average ages in years for the assessment points were 9, 11, and 15. We considered within time point age variation (see Table 1), but it did not influence any findings so those analyses were not reported. Attrition and other factors, such as not completing all of the assessments at each time point, led to some variation in sample sizes across the groups. Procedures for handling incomplete data will be addressed in the Results section.

Characteristics of the participants are shown in Table 1. The sample was nearly evenly split by sex. SES, measured by mother's level of education, was reported at the beginning of the overall study when the child was 1 month of age. Average education level for the entire study sample was 14.4 years. The sample included approximately 80% of participants identifying as European Americans, 12% as African American, and 6% as Hispanic. Asian Americans and Native Americans accounted for approximately 2% of the total sample.

### 2.2. Measurement instruments

#### 2.2.1. Woodcock–Johnson measures

Selected tests from the Woodcock–Johnson Psychoeducational Battery — Revised (WJ-R; Woodcock & Johnson, 1989) were used in this research. The WJ-R includes measures of intelligence and academic achievement. These measures were administered during lab visits at each of the participating sites. In this study, the Picture Vocabulary, Passage Comprehension, and Letter-Word Identification tests were

**Table 1**
Sample characteristics.

| Characteristic | *N* | Percentage |
|---|---|---|
| Sex | | |
| Male | 543 | 50.3 |
| Female | 536 | 49.7 |
| Maternal education | | |
| <High school | 89 | 8.2 |
| High school/GED | 217 | 20.1 |
| Some college | 258 | 33.2 |
| Bachelor's degree | 244 | 22.6 |
| Beyond bachelor's degree | 171 | 15.8 |
| Race/ethnicity[a] | | |
| Caucasian | 877 | 81.3 |
| African-American | 130 | 12.0 |
| Asian/Pacific Islander | 16 | 1.5 |
| Native American | 4 | 0.4 |
| Other | 52 | 4.8 |
| Hispanic | 65 | 6.0 |

*Note.* All specific responses reporting <High School education (e.g., completed ninth grade and completed tenth grade) were combined; all specific responses Beyond Bachelor's (e.g., M.A. and PhD) were also combined.

[a] Total exceeds 100% due to some participants selecting more than one category.

used as measures of Gc, reading comprehension (RC), and sight word identification, respectively. The psychometric properties of these well-known tests are excellent (McGrew, Werder, & Woodcock, 1991).

Picture vocabulary required participants to expressively label a pictured object following brief oral prompts. The scores demonstrated adequate internal consistency at each time point in the current sample (Cronbach's $\alpha = .72–.81$). Vocabulary tests are often considered some of the best measures of intelligence, and of Gc in particular (Horn & McArdle, 2007). Consistent with Spearman's (1927) idea of educing correlates, vocabulary development depends greatly on the ability to educe meaning from context, and "the acquisition of vocabulary is not as much a matter of learning and memory as it is of generalization, discrimination, eduction, and inference" (Jensen, 1980, p. 146). Thus, although it was a single measure of vocabulary, this measure was an excellent Gc proxy, and had shown to load strongly on a Gc factor in factor analysis (Bickley, Keith, & Wolfle, 1995; Woodcock, 1990).

The Passage Comprehension test employs a cloze procedure to measure the ability to extract meaning from connected text. Participants were required to read a short passage and utilize the context of the passage to supply a word that was missing from the passage. The scores from this test also demonstrated high internal consistency at each time point (Cronbach's $\alpha = .81–.83$) in the current sample.

The Letter-Word Identification test from the WJ-R, administered in 3rd grade, was used as a measure of basic sight word identification (Cronbach's $\alpha = .90$). Participants were required to read and identify letters and words, independent of context. In contrast to the scores from the other two WJ-R tests, these scores were mean centered.

Given the nature of the analysis, *W*-scores were modeled. *W*-scores are Rasch-based scores that contain properties ideal for longitudinal analysis. These scores are on a common metric and are in equal interval units across age and tests. The mean is centered at 500, which is an approximation of the average score of a 10 year old (Mather & Woodcock, 2001).

#### *2.2.2. Bayley Scales of Infant Development-II*

Infant intellectual functioning, which we will refer to as *early childhood intelligence*, was measured with the Bayley Scales of Infant Development — Revised Mental Developmental Index (BSID-II; Bayley, 1993). The measure was administered at each participating site when the child was 24-months-old. The BSID-II includes several tasks designed to assess a young child's response to stimuli, problem solving, and language development. Scores are on a standard scale ($M = 100$, $SD = 15$), although scores were mean centered for the analyses.

Internal consistency reliability estimates for scores from the BSID-II have ranged from .78 to .93 (Bradley-Johnson, 2001). An estimate of internal consistency reliability for scores from those in this sample with complete data and who were administered the same range of items ($n = 869$) was within that range ($\alpha = .84$). The BSID-II has demonstrated concurrent-related validity with other early scales of cognitive abilities (Bradley-Johnson, 2001), including the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1989). Scores from the BSID-II administered at 30 months are predictive of nonverbal cognition during the preschool years and have demonstrated continuity with other intellectual and verbal abilities (Blaga et al., 2009). In addition, scores have shown to be stable over time for children 24 and 36 months ($r = .91$) and have demonstrated high inter-rater reliability ($r = .96$; Bayley, 1993; Gagnon & Nagle, 2000).

#### *2.2.3. Title recognition task*

The Title Recognition Task (TRT; Cunningham & Stanovich, 1991) was included as a measure of relative reading volume. In 5th grade, children were required to identify real book titles from a list of actual and imagined titles. The task was modified from the original for the SECCYD study, reducing the total number of titles from 100 to 50. The measure was administered by research staff during home visits. Scores were computed by calculating the difference between the proportion of correct titles marked and proportion of imagined titles marked. Higher scores indicated higher levels of title recognition and thus, relatively higher reading volume. Possible scores ranged from −1 to 1, although values in this sample ranged from −.36 to .80. Similar recognition tasks have demonstrated convergent validity with other measures of print exposure, as well as adequate internal consistency reliability ($\alpha = .78–.80$; Allen, Cipielewski, & Stanovich, 1992). Scores from the TRT have demonstrated consistency in effects across time, with similar predictive validity across 4th through 6th grades for measures of spelling, vocabulary, and general knowledge (Cunningham & Stanovich, 1991).

### 2.3. Model specification

#### 2.3.1. Bivariate dual change score model

The first step was to establish a bivariate dual change score model for Gc and reading comprehension (RC). Passage Comprehension and Picture Vocabulary scores were collected on three occasions (grades 3, 5, and age 15), and each of those occasions is represented by the rectangles (Fig. 1). To ease in the explanation of this model of a dynamical system, the model for Gc scores, located in the upper portion of Fig. 1, will be described in detail. The same description applies to the RC scores, located in the lower portion of Fig. 1. To ease interpretation, grade levels were used for latent variable descriptors (e.g., age 15 scores were labeled as Gc 9).

First, observed Gc scores at each time point (appearing as rectangles in the figures) were decomposed into true scores (ovals labeled Gc3, Gc5…) and separated from measurement error (ovals labeled E1). Factor loadings were constrained to be invariant over time. Measurement error was held constant and

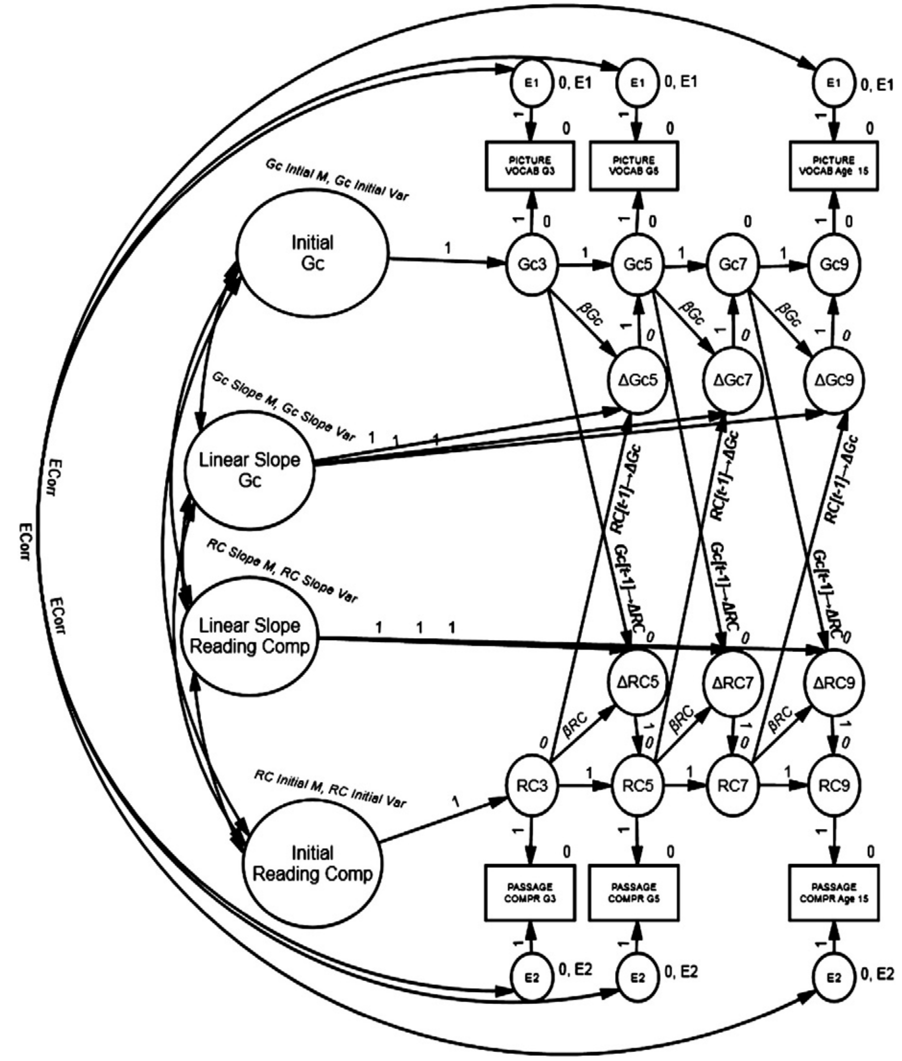


**Fig. 1.** Bivariate dual change score model for Gc and reading comprehension.

uncorrelated over time. Next, latent change scores were specified as second-order factors (ovals labeled $\Delta Gc_5$, $\Delta Gc_7$, $\Delta Gc_9$). Two internal forces directly exerted their influence on the latent change scores at each time point: one constant (labeled Linear Slope Gc) and one autoproportion or autoregressive effect ($\beta_{Gc}$). As shown in Fig. 1, the first direct influence on $\Delta Gc_5$ exited from a third-order factor, with each of its paths fixed to 1. The latent mean of the third-order factor, labeled Linear Slope Gc, represented the average constant slope (or internal natural growth), with its variance representing individual differences in those slopes. The second direct influence, the proportional effect ($\beta_{Gc}$) on Grade 5 latent change scores ($\Delta Gc_5$), exited Gc3 and entered $\Delta Gc_5$. That is, $\Delta Gc_5$ were regressed on previous true scores. These paths were constrained equal across time. So, $\Delta Gc_5$ was a function of two types of internal change, and hence the name *dual change*: a constant amount (Gc Linear Slope) and a proportional amount ($\beta_{Gc}$).

In a bivariate dual change score model, such as shown with Gc and RC in Fig. 1, however, there is also an external force influencing these changes (McArdle, 2009). These forces are referred to as *coupling parameters* (e.g., paths labeled $RC_{t-1} \rightarrow \Delta Gc$). These are direct, time-dependent effects of one variable on change in the other variable (e.g., the direct influence of prior RC on change in Gc). Therefore, in the bivariate dual change score model, $\Delta Gc_5$ was a direct function of its latent slope (Gc Linear Slope), its autoproportion ($\beta_{Gc}$) – as described previously – and the level of RC in 3rd grade ($RC_{t-1} \rightarrow \Delta Gc_5$).

A few additional comments about this model are needed (Fig. 1). Note the Initial Gc factor in the top left of the model and its path entering Gc3. The Initial Gc factor mean represented the mean of Gc scores in 3rd grade, and its variance represented individual differences about that mean. In addition, measurement errors for Gc and RC observed variables (i.e., ovals labeled E1 and E2, respectively) were correlated within each time point and fixed to be equal over time (ECorr). Last, note the latent variables representing the observed scores for Grade 7 (Gc7 and Rc7). These "node variables" were included because data were not collected in Grade 7. Including these node variables allowed for equal interval measurements (approximately 2 years). Thus, time was appropriately "stretched out" (Ferrer et al., 2004).

To summarize, this model included two fixed effects representing initial Gc and RC levels (Initial factor means), with individual differences in those levels (Initial factor variances), two fixed effects representing a constant linear slope for Gc and RC (Linear Slope factor means), individual differences in each of those slopes (Linear Slope factor variances), six covariances between Initial and Linear Slope factors, two estimates of proportional change ($\beta_{Gc}$; $\beta_{RC}$), two coupling parameters ($RC_{t-1} \rightarrow \Delta Gc$; $Gc_{t-1} \rightarrow \Delta RC$), a single estimate for each construct's error variance (E1; E2), and a single estimate representing within time correlations between error variances (ECorr). Change in these models was cumulative; $\Delta Gc_5$ did not go backward in time or directly influence variables later in time. Indirectly, however, $\Delta Gc_5$ influenced $Gc_7$, $Gc_9$, each later observed score, and all later changes. All previous factors had an influence directly or indirectly on all later measurements so that the changes were accumulated over time. Moreover, although the models were specified as linear, with coefficients specified as constants, these constants were multiplied by previous change scores so the average trajectories over time may be nonlinear (McArdle, 2009). In fact both average trajectories were expected to be nonlinear, specifically increasing but decelerating over time.

*2.3.2. Bivariate dual change score model with covariates*

Background variables and covariates were included in a conditional bivariate dual change score model. All of these variables except for sex were mean centered. The Gc and RC Initial and Slope factors (from Fig. 1) were regressed directly on the three background variables: SES, sex, and early childhood intelligence. Basic sight word recognition and relative reading volume scores were also included. These two variables were regressed on the background variables, with the Initial and Slope factors also regressed directly on these variables. Therefore, the three background variables had total, direct, and indirect effects on the Initial and Slope factors. Sight word recognition and relative reading volume had direct effects, or the total effects equaled the direct effects, and were considered potential mediators. Each of these variables was time invariant.

## *2.4. Model evaluation*

Model fit indexes were used to evaluate and compare models. Fit of individual models were evaluated using the chi squared ($\chi^2$) test-of-fit statistic, root mean square error of approximation (RMSEA; Steiger & Lind, 1980), and comparative fit index (CFI; Bentler, 1990). Comparisons of nested models were made with the likelihood ratio test after Satorra–Bentler corrections (Satorra & Bentler, 1994) were applied. These corrections were necessary for the MLR estimator because the difference between nested models using that estimator is not distributed as $\chi^2$. Bayesian information criteria were also reported for model comparisons (BIC; Schwarz, 1978). Lower BIC values indicated better fitting models.

## 3. Results

### 3.1. Descriptive statistics

Descriptive statistics are shown in Table 2. When examining mean values, it is apparent that Gc and RC *W*-scores increased over time. There was increased dispersion of Gc *W*-scores at the last time point, whereas there was decreased dispersion of RC *W*-scores after the first time point. The correlations between variables within constructs showed a simplex pattern, with those closer in time correlated more strongly. The background variables and covariates, except for sex, had a mean near zero because they were mean-centered prior to the analysis.

Absolute univariate skewness values that approach or exceed 2 and absolute kurtosis values that approach or exceed 7 may result in problems with non-normality (Curran, West, & Finch, 1996). Absolute univariate kurtosis values were less than 4 and skewness values were less than 2 for the Gc and RC variables after one case was deleted. Scores from this case produced large differences in the distributional assumptions, causing the kurtosis value for Grade 5 RC to be very large (>8). One participant had minimal raw scores on the measure across time, which indicated an inadequate test floor for this person. When this case was deleted, all of the univariate distributional assumptions were within appropriate limits.

Although univariate skewness and kurtosis were acceptable, we used the MLR estimator, which is robust to departures from normality, within M*plus* for our models (Muthén & Muthén, 1998–2010). High levels of kurtosis, leptokurtotic distributions in particular, have been found to result in inflated $\chi^2$ values and smaller standard errors. The latter results in a risk of Type I error for specific parameters, whereas the former leads to a greater likelihood of rejecting the overall model. Maximum likelihood (ML) estimation was also used. Results were compared to those obtained from the MLR estimator. Standard errors were smaller with the ML estimator; thus, a couple parameters were statistically significant using the ML estimator, whereas they were not with the MLR estimator. There was little difference in $\chi^2$ values. The minimal differences in the statistically significant findings did not influence substantively important conclusions. Because there were some standard error differences, however, we report results from the MLR estimator because the differences in standard errors may have indicated some multivariate

**Table 2**
Sample statistics and correlations between Gc, reading comprehension, and selected covariates.

| Measure | Gc3 | Gc5 | Gc9 | RC3 | RC5 | RC9 | SES | BSID-II | LW3 | RV5 | Age G3 | Age G5 | Age G9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gc3 | – | | | | | | | | | | | | |
| Gc5 | **.79** | – | | | | | | | | | | | |
| Gc9 | **.74** | **.81** | – | | | | | | | | | | |
| RC3 | .64 | .63 | .59 | – | | | | | | | | | |
| RC5 | .60 | .63 | .63 | **.77** | **–** | | | | | | | | |
| RC9 | .61 | .66 | .70 | **.67** | **.70** | – | | | | | | | |
| SES | .41 | .42 | .42 | .37 | .39 | .40 | – | | | | | | |
| BSID–II | .41 | .43 | .40 | .44 | .42 | .42 | .37 | – | | | | | |
| LW3 | .55 | .53 | .50 | .79 | .69 | .56 | .33 | .33 | – | | | | |
| RV5 | .51 | .54 | .56 | .52 | .52 | .55 | .44 | .37 | .52 | – | | | |
| *N* | 1013 | 991 | 888 | 1010 | 991 | 886 | 1079 | 1008 | 932 | 561 | 1013 | 992 | 891 |
| *M* | 496.98 | 505.91 | 518.66 | 495.38 | 505.29 | 520.41 | 0.00 | 0.00 | 0.00 | 0.00 | 8.99 | 10.99 | 15.03 |
| *SD* | 11.43 | 11.91 | 13.05 | 14.25 | 12.24 | 12.41 | 2.45 | 14.59 | 18.30 | 0.22 | 0.31 | 0.33 | 0.14 |
| Min | 440 | 440 | 451 | 412 | 425 | 446 | −7.42 | −42.62 | −127.30 | −0.59 | 8.25 | 10.25 | 14.67 |
| Max | 531 | 542 | 563 | 527 | 542 | 554 | 6.58 | 57.38 | 41.75 | 0.57 | 10.17 | 12.33 | 15.92 |

*Note.* Bold = within construct correlations. Gc = picture vocabulary; RC = passage comprehension; SES = maternal education; BSID = Bayley Scale of Infant Development-II; LW = letter-word identification; RV = reading volume. Min = minimum value; Max = maximum value.
All correlations were statistically significant at $p<.05$.

nonnormality. It is important to note that our choice of estimator did not affect the parameter estimates; it only resulted in corrected standard errors and model $\chi^2$.

### 3.2. Incomplete data

There were some missing data, with fewer cases available at later times (Table 2). Little's (1988) missing completely at random (MCAR) test revealed that the null MCAR hypothesis could not be rejected, $\chi^2$ (54) = 69.65, $p$ = .07. Rather than delete cases with incomplete data, which is acceptable under the MCAR assumption, we used all available cases. The MLR estimator is appropriate for nonnormal and missing data under a variant of the less strict assumption of missing at random (MAR; Yuan & Bentler, 2000).

Further analysis was performed when the covariates were included. For example, the relative reading volume variable included estimates for 561 cases. We ran the analysis with and without the mediating variables. No substantive changes in total effects from the background variables were noted, with almost identical standardized and unstandardized estimates. Therefore, all of the results are based on information from all available cases, whether the data were incomplete or not, under the assumption MAR.

### 3.3. Bivariate dual change score models

The bivariate dual change score model with full couplings between the variables fit very well, $\chi^2$ (6) = 11.33, $p$ = .08, CFI = .999, RMSEA = .029, BIC = 41379.5. The two parameters of greatest interest were the couplings. The RC coupling parameter (prior Gc to RC) was positive and statistically significant ($Gc_{t-1} \rightarrow \Delta RC = 0.48$, $p < .05$). The Gc coupling (prior RC to Gc), however, was negative, tiny, and not statistically significant ($RC_{t-1} \rightarrow \Delta Gc = -0.03$, $p = .80$), indicating that prior Gc led to positive change in RC, but not vice versa.

Two additional models were estimated, each nested within the full coupling model: a model with the path from Gc to RC fixed to zero ($Gc_{t-1} \rightarrow \Delta RC = 0$) and a model with the path from RC to Gc fixed to 0 ($RC_{t-1} \rightarrow \Delta Gc = 0$). The $\chi^2$ difference tests, after Satorra–Bentler corrections were applied, were also consistent with the statistical significance of the parameters obtained from the full coupling model. The $RC_{t-1} \rightarrow \Delta Gc$ path was fixed to zero, ($\chi^2$ [7] = 11.09, CFI = .999, RMSEA = .023, BIC = 41372.6), and there was no degradation in model fit, Satorra–Bentler $\Delta\chi^2$ (1) = 0.71, $p$ = .79. There was a statistically significant degradation in model fit (Satorra–Bentler $\Delta\chi^2$ [1] = 4.95, $p < .05$) when $Gc_{t-1} \rightarrow \Delta RC$ was fixed to zero, $\chi^2$ (7) = 18.67, CFI = .997, RMSEA = .039, BIC = 41382.8. In addition, the BIC values indicated that the model with $Gc_{t-1} \rightarrow \Delta RC$ estimated and $RC_{t-1} \rightarrow \Delta Gc$ fixed to zero provided the best fit. Taken together, these findings answered our first research question: $Gc_{t-1} \rightarrow \Delta RC$ was statistically significant, but not vice versa.

Remaining estimates from the bivariate model with $RC_{t-1} \rightarrow \Delta Gc = 0$ are shown in Table 3.[4] As expected, Initial factor means, or the average starting point for those students in 3rd grade, for both Gc ($M$ = 496.88) and RC ($M$ = 495.32) were statistically significant, as was the variance associated with those factors. Correlations among the Initial and Slope factors are also shown in Table 3. The 3rd-grade Gc and RC scores correlated strongly ($r$ = .80), 3rd-grade Gc correlated positively with constant change in Gc ($r$ = .85), and 3rd-grade RC correlated positively with Gc ($r$ = .66) and RC ($r$ = .49) slopes. The slopes also correlated positively ($r$ = .26), although this correlation was not statistically significant.

The estimates related to biannual changes in Gc and RC, which were described via Linear Slope factor means, autoproportions, and couplings were of great interest. Linear Slope factor means for both variables ($M_{Gc} = 121.48$ and $M_{RC} = 70.78$) were statistically significant and positive, with statistically significant variation ($SD_{Gc} = 3.39$ and $SD_{RC} = 3.39$) about those means (see Table 3). Autoproportions for both variables ($\beta_{Gc} = -0.226$; $\beta_{RC} = -0.599$) were statistically significant, negative, and ergodic; internal growth in both variables was constrained or slowed by higher previous levels of that variable. In general, and without yet interpreting the couplings, on average there was positive, constrained growth. The

[4] The estimates obtained from this model and the initial bivariate model were very similar, so we presented the findings from this more parsimonious model.

**Table 3**
Parameter estimates for bivariate the dual change score model.

| Parameter | Gc | | Reading comprehension |
|---|---|---|---|
| Fixed | | | |
| Initial mean | 496.88 | | 495.32 |
| Slope mean | 121.48 | | 70.79 |
| Proportion | −0.23 | | −0.60 |
| Coupling | 0.00[a] | | 0.47 |
| Random | | | |
| Initial variance | 102.93 | | 165.02 |
| Slope variance | 11.53 | | 11.49 |
| Error variance | 27.66 | | 38.15 |
| Correlations | | | |
| Initial Gc ↔ initial RC | | .80 | |
| Initial Gc ↔ slope Gc | | .85 | |
| Initial Gc ↔ slope RC | | .00[b] | |
| Initial RC ↔ slope RC | | .49 | |
| Initial RC ↔ slope Gc | | .66 | |
| Slope RC ↔ slope Gc | | .26[b] | |
| Within-time residuals | | .02 | |

*Note.* Values are statistically significant unless otherwise noted.
[a] Parameters were fixed to 0.
[b] Results are *not* statistically significant at $p<.05$.

interpretation is incomplete, however, without interpretation of the statistically significant coupling parameter ($Gc_{t-1} \rightarrow \Delta RC = 0.474$, $p<.01$).[5]

Biannual change equations for Gc and RC were expressed as

$$\Delta G[t] = 121.48 \pm 3.39 - .226Gc_{t-1}$$
$$\Delta R[t] = 70.78 \pm 3.39 - .599RC_{t-1} + .474Gc_{t-1}.$$

Gc increased biannually by 121.48 *W*-scale points ($SD = 3.39$) as indicated by the Gc slope, but the increase was slowed by previous Gc, as indicated by the autoproportion (−.226). Prior RC was not directly related to changes in Gc. Alternatively, RC increased biannually by 70.78 *W*-scale points ($SD = 3.39$). RC increases were slowed by higher prior levels of RC (−.599) but accelerated by higher levels of prior Gc (.474). Gc was a leading indicator of RC growth but not vice versa (See Fig. 2).

The development of Gc and RC may be further explored by examining plots of model implied means. The development of Gc from 3rd to 9th grades was plotted at different initial levels of Gc: initial 3rd-grade estimates were at the mean, one standard deviation above the mean, and one standard deviation below the mean (Fig. 3). These trajectories represent internal growth, which included information from the Initial level, Linear Slope, and autoproportion parameters. In general, Gc shows an increasing yet decelerating form.

To understand the longitudinal development of RC, the coupling parameter ($Gc_{t-1} \rightarrow \Delta RC$) needed to be considered. In Fig. 4, RC development is a function of the model (i.e., dynamic system) at different initial RC and Gc sample means. There are three panels within Fig. 4. Initial RC values were one standard deviation above the mean in Panel A, at the mean in Panel B, and one standard deviation below the mean in Panel C. Within each Panel, RC trajectories were plotted as a function of the dynamic system at different initial levels of Gc. Different initial levels of Gc had noticeable influences on the developmental trajectories of RC. Higher levels of initial Gc resulted in steeper RC growth within each initial level of RC.

Last, we again plotted RC estimates (see Fig. 5); however, different initial levels of RC were plotted with the initial Gc value held constant (i.e., the Gc mean at the initial time point) across RC levels. Initial estimates of RC were one standard deviation above the mean, at the mean, and one standard deviation below the mean. As shown in Fig. 5, despite different initial RC estimates, similar initial levels of Gc had an

[5] The coupling is slightly smaller than the coupling in the initial bivariate model.

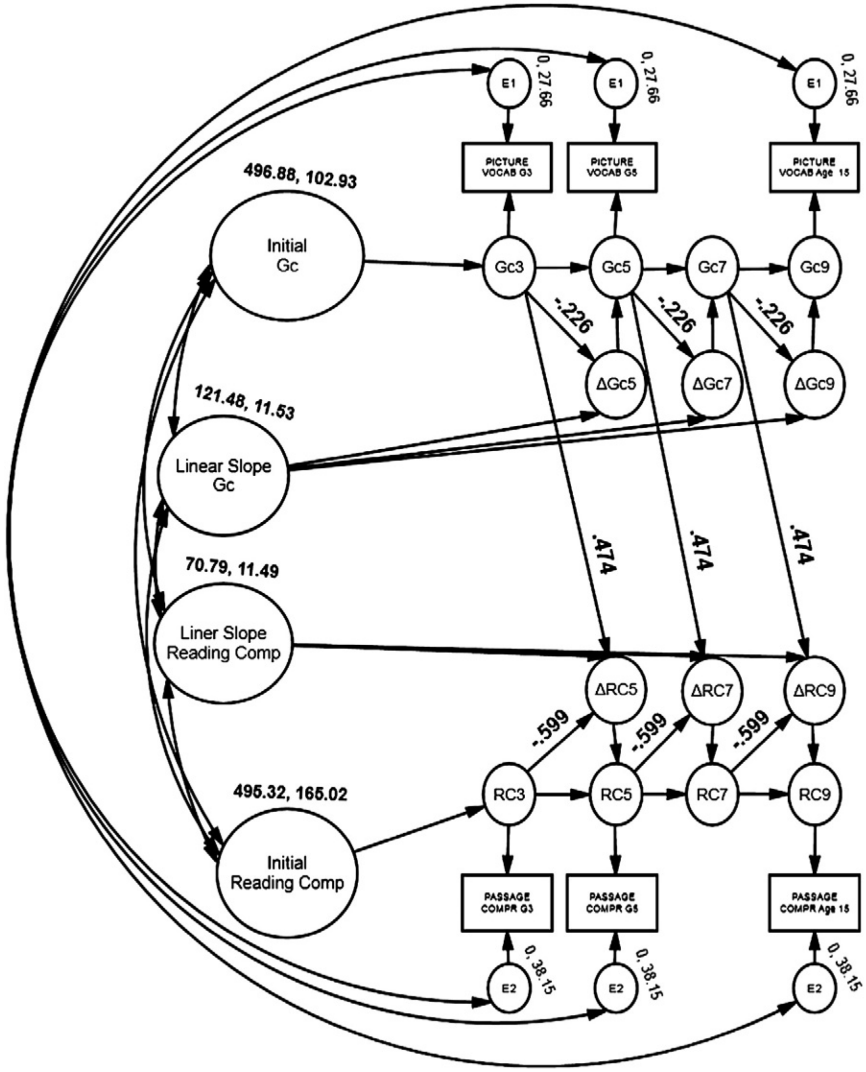


**Fig. 2.** Bivariate dual change score model for Gc and reading comprehension with relevant model parameter estimates.

equalizing influence on the RC scores over time. In other words, the 3rd-grade reading comprehension gap was reduced substantially in 9th grade.

### 3.4. Bivariate dual change score with time-invariant covariates

Time-invariant covariates were included as explanatory variables in a conditional bivariate dual change score full coupling model. Model fit was excellent, $\chi^2$ (16) = 17.21, $p$ = .37, RMSEA = .01, CFI = 1.00. Direct, total indirect, and total effects for SES, sex, and early childhood intelligence on the Initial and Slope factors are shown in Table 4, as are the direct (i.e., total) effects for sight word recognition and relative reading volume.

First, inclusion of the covariates had no noticeable effects on the autoproportions or couplings, suggesting no alterations to the dynamical system. These variables, however, influenced initial levels and slopes of these constructs. Boys scored about 3 *W*-score points higher on Gc in 3rd grade (see Table 4).

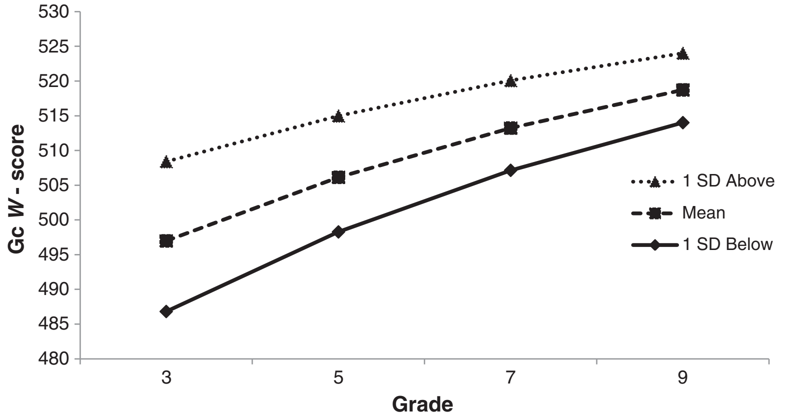


**Fig. 3.** Expected average latent trajectories for Gc. Gc model implied *W*-score means are plotted based on the change equation, with the trajectories plotted for different 3rd grade Gc *W*-score sample means (1 *SD* below the *M*, *M*, and 1 *SD* above *M*).

Sight word recognition and relative reading volume did not mediate that effect. Sex also had a statistically significant effect on the internal change in Gc, with boys showing steeper increases. There were no sex differences in 3rd-grade reading comprehension, although girls showed steeper growth in RC over time.

SES had a positive and statistically significant total effect on 3rd-grade Gc. More than half of the SES effect (59%) was mediated by sight word recognition and relative reading volume. SES also had a statistically significant total indirect effect on 3rd-grade RC. Although basic reading and reading volume mediated most of this influence (81%), a small yet statistically significant direct effect remained. SES had a direct influence on Gc slopes, with 62% of the total effect of SES on Gc slopes mediated by sight word recognition and relative reading volume. SES had a small yet statistically significant total indirect effect on RC slopes.

Early childhood intelligence exerted positive and statistically significant total effects on 3rd-grade Gc and RC levels, as well as the slopes. Early childhood intelligence had relatively larger total effects on Initial Gc and RC than did SES. About 44% of the total effect on 3rd-grade Gc was mediated by basic reading and reading volume, but a statistically significant direct effect remained (see Table 4). Similarly, although sight word recognition and relative reading volume mediated about 64% of the total effect on 3rd-grade RC, a statistically significant direct effect remained. Early childhood intelligence had statistically significant and positive indirect and direct effects on linear growth in Gc, but the total effect on RC internal growth was completely mediated by sight word recognition and relative reading volume. An interpretation of total effects indicates that children with higher intelligence scores at 24 months of age demonstrated higher levels of Gc and RC in 3rd grade as well as steeper linear growth in these constructs through 9th grade, holding SES and sex constant.

Last, the two mediating variables also exerted statistically significant direct effects on initial levels of Gc and RC (see Table 4). Those with higher word reading scores and who read more also had higher 3rd-grade Gc scores. Both basic word recognition skills and relative reading volume were associated with linear growth in Gc scores but not RC scores. Taken together, the covariates accounted for 82% and 20% of the variance in the Initial and Slope RC factors, respectively. They accounted for 57% and 55% of the variance in the Initial and Slope Gc factors, respectively.

## 4. Discussion

The development of Gc and reading comprehension in children from grade 3 through grade 9 was investigated. Of specific interest were the dynamic longitudinal interrelations between Gc and reading comprehension. Bivariate dual latent change score models (McArdle, 2009), with and without time-invariant covariates, were applied to Rasch-based *W*-scores from the WJ-R Passage Comprehension and Picture Vocabulary tests. In the first set of findings, Gc showed positive yet decelerating growth

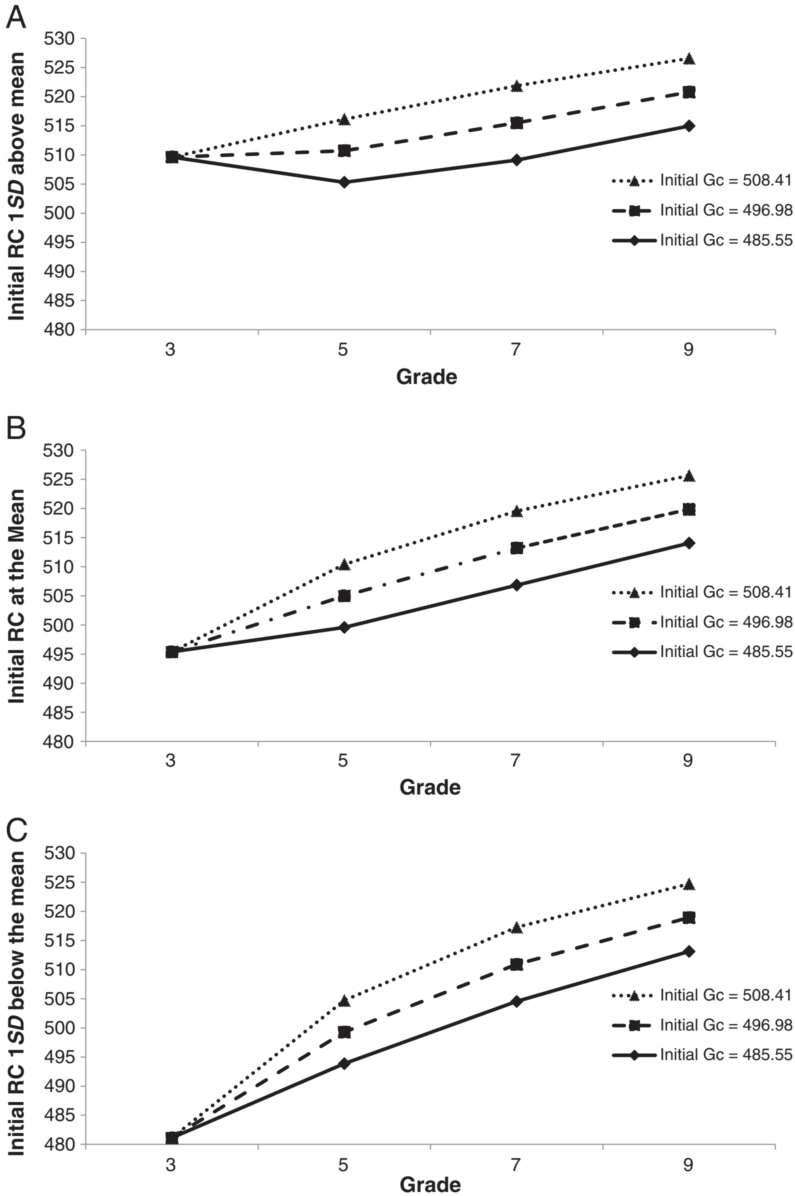


**Fig. 4.** Expected average latent trajectories for RC based on the bivariate dual change score equation. Panel A includes 3rd-grade RC *W*-scores that were 1 *SD* above *M*, with RC as a function of different 3rd-grade Gc *W*-score sample means (1 *SD* below *M*, *M*, and 1 *SD* above *M*). Panel B includes 3rd-grade RC *W*-scores that were at the sample mean, with RC as a function of different 3rd-grade Gc *W*-score sample means. Panel C includes 3rd-grade RC *W*-scores that were 1 *SD* below *M*, with RC as a function of different 3rd-grade Gc *W*-score sample means.

through 9th grade. Reading comprehension also showed positive, decelerating growth, yet growth in reading comprehension was accelerated by higher levels of prior Gc. That is, there was a unidirectional dynamic longitudinal relation between constructs, with prior Gc as a positive leading indicator of growth in reading comprehension, not vice versa.

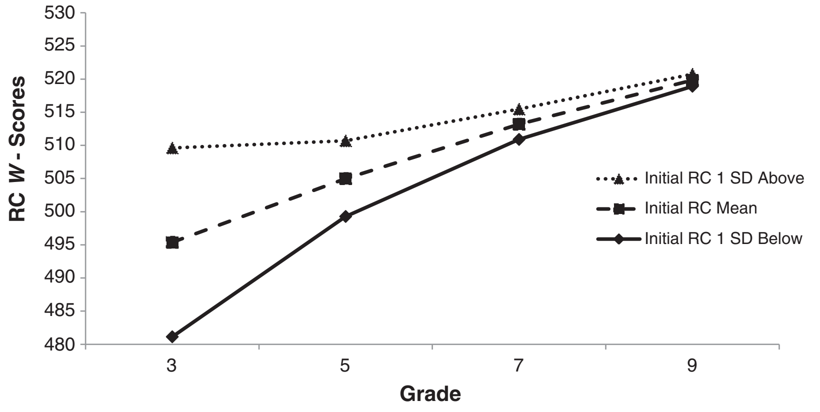


**Fig. 5.** Expected latent RC trajectories for different 3rd-grade RC *W*-score sample means, with the same 3rd-grade Gc *W*-score sample mean. The 3rd-grade *W*-score sample means of RC were at 1 $SD<M$, $M$, and 1 $SD>M$. The 3rd-grade Gc scores were at the sample mean (*W*-score $M=496.98$).

A second set of findings involved the covariates. In terms of total effects, SES, 3rd-grade sight word reading, and relative reading volume had positive effects (a) on Gc and reading comprehension in 3rd grade and (b) on Gc slopes. Children who had higher intelligence scores when they were 24 months-old had higher 3rd-grade Gc and reading comprehension scores and steeper Gc and reading comprehension slopes. Boys had higher 3rd-grade Gc scores and steeper Gc slopes. Girls had steeper reading comprehension slopes. The covariates did not alter the dynamic parameters.

### 4.1. Relations with prior research

One purpose of this study was to replicate previous findings from Ferrer et al. (2007, 2010). Our findings diverged in some ways from those two studies. In one study, Ferrer et al. (2010) found a small,

**Table 4**
Total, total indirect, and direct effects of the covariates on Gc and reading comprehension initial and slope factors.

| | Initial Gc | | | Gc slope | | | Initial RC | | | RC slope | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Covariate | Total | Total indirect | Direct | Total | Total indirect | Direct | Total | Total indirect | Direct | Total | Total indirect | Direct |
| Sex | **−.14** | **.01** | **−.15** | **−.13** | **.01** | **−.14** | **.01** | **−.02** | **.03** | **.18** | **.00** | **.18** |
| | −2.79$^{*}$ | 0.17 | −2.95$^{*}$ | −0.82$^{*}$ | 0.11$^{*}$ | −0.94$^{*}$ | 0.29 | −0.38 | 0.66 | 1.21$^{*}$ | 0.00 | 1.21 |
| | (0.59) | (0.34) | (0.54) | (0.41) | (0.13) | (0.43) | (0.74) | (0.53) | (0.54) | (0.60) | (0.10) | (0.65) |
| SES | **.32** | **.19** | **.13** | **.34** | **.21** | **.13** | **.27** | **.22** | **.05** | **.11** | **.12** | **−.01** |
| | 1.34$^{*}$ | 0.79$^{*}$ | 0.55$^{*}$ | 0.46$^{*}$ | 0.28$^{*}$ | 0.18$^{*}$ | 1.43$^{*}$ | 1.16$^{*}$ | 0.27$^{*}$ | 0.15 | 0.16$^{*}$ | −0.01 |
| | (0.14) | (0.09) | (0.14) | (0.06) | (0.04) | (0.06) | (0.16) | (0.14) | (0.13) | (0.09) | (0.05) | (0.10) |
| BSID | **.38** | **.18** | **.20** | **.32** | **.18** | **.14** | **.38** | **.24** | **.14** | **.16** | **.12** | **.04** |
| | 0.27$^{*}$ | 0.12$^{*}$ | 0.14$^{*}$ | 0.07$^{*}$ | 0.04$^{*}$ | 0.03$^{*}$ | 0.33$^{*}$ | 0.21$^{*}$ | 0.12$^{*}$ | 0.04$^{*}$ | 0.03$^{*}$ | 0.01 |
| | (0.03) | (0.02) | (0.02) | (0.01) | (0.01) | (0.01) | (0.03) | (0.03) | (0.02) | (0.01) | (0.01) | (0.02) |
| LW3 | **.36** | _ | **.36** | **.28** | _ | **.28** | **.75** | _ | **.75** | **.29** | _ | **.29** |
| | 0.20$^{*}$ | | 0.20$^{*}$ | 0.05$^{*}$ | | 0.05$^{*}$ | 0.52$^{*}$ | | 0.52$^{*}$ | 0.05 | | 0.05 |
| | (0.02) | | (0.02) | (0.02) | | (0.02) | (0.02) | | (0.02) | (0.03) | | (0.03) |
| RV5 | **.29** | _ | **.29** | **.39** | _ | **.39** | **.11** | _ | **.11** | **.14** | _ | **.14** |
| | 12.93$^{*}$ | | 12.93$^{*}$ | 5.68$^{*}$ | | 5.68$^{*}$ | 6.20$^{*}$ | | 6.20$^{*}$ | 2.01 | | 2.01 |
| | (1.87) | | (1.87) | (1.14) | | (1.14) | (1.86) | | (1.86) | (1.70) | | (1.70) |

*Note.* Bold numbers are standardized estimates and nonbold numbers are unstandardized. Standard errors are in parentheses. Gc = picture vocabulary; RC = passage comprehension; SES = maternal education; BSID = Bayley Scale of Infant Development-II; LW = letter-word identification; RV = reading volume.
$^{*}$ $p<.05$.

unidirectional relation from FSIQ to reading in individuals with dyslexia and positive mutually beneficial relations between FSIQ and reading in a sample of those without dyslexia. In another study, Ferrer et al. (2007) found positive, mutually beneficial relations between Verbal IQ (i.e., Gc) and reading comprehension. We found a unidirectional dynamic link from Gc to reading comprehension.

There are several possible reasons for divergent findings. We will discuss two. First, Ferrer et al. (2007, 2010) used measures of reading and intelligence that were collected in 1st grade. Basic reading and reading comprehension scores were combined into a reading composite, and models allowing for differential coupling effects across grades were estimated. The largest dynamic interrelations between Verbal IQ and reading occurred before 4th grade. In our study, data collection for reading comprehension and Gc began in 3rd grade, potentially negating the opportunity to observe important prior influences. Second, different measures were used across the three studies. In Ferrer et al. (2007) study, prior reading comprehension may have influenced changes in more specific constructs included in a Verbal IQ composite score, which were unrelated to Gc. For example, the Verbal IQ from the WISC-R (Wechsler, 1974) was confounded with a measure of Arithmetic. It is unclear what Arithmetic measures, but it typically has the highest *g* loadings of any subtest, and it is generally considered a cognitively complex task. Some have found that it may be an indicator of fluid reasoning (e.g., Keith, Fine, Taub, Reynolds, & Kranzler, 2006), and a mutually beneficial relation between Gf and academic achievement has been demonstrated in prior research (Ferrer & McArdle, 2004). Similarly, in the Ferrer and colleagues (2010) study, prior reading may have influenced other specific aspects of intelligence in general that are summarized in a FSIQ score in a sample without dyslexia. The converse is also true, namely related to the construct underrepresentation of Gc in our study, in which Gc was measured by one indicator. Perhaps it was the narrow construct of vocabulary knowledge that was not influenced by prior reading comprehension. Similarly, we modeled reading comprehension, whereas some of the previous findings have been related to reading in general. Future research examining dynamic changes involving more broad and narrow aspects of intelligence and academic achievement from well-defined latent variables will be enlightening.

### *4.2. General findings and theories*

Although our findings diverged somewhat from Ferrer et al. (2007, 2010) prior research, they were consistent with findings from other studies. For example, IQ has been found to be a leading indicator of academic achievement but not vice versa when using cross-lag panel designs (Watkins, Lei, & Canivez, 2007). We will discuss in the paragraphs that follow how our findings may be interpreted within the context of three general findings and theories.

First, verbal reasoning or proficiency has been found to exert an important influence on reading comprehension in children (Keith, 1999; Tilstra et al., 2009). Here, prior levels of Gc indeed exerted an increased positive change in reading comprehension. Perhaps maturation or cortical changes improved the absolute capacity to educe correlates or make inferences with verbal information, which in turn increased the readiness to acquire more vocabulary knowledge. These processes may have been reflected first in measures of vocabulary knowledge but take longer to manifest in measures of reading comprehension because reading comprehension also requires learning how to apply that knowledge to print. Hence, vocabulary knowledge is a leading indicator of reading comprehension.

Second, our findings seem consistent with the simple view of reading (Hoover & Gough, 1990), in which language comprehension and basic word reading skills underlie reading comprehension. Language comprehension may be considered a Gc proxy. When basic word reading skills are mastered, Gc abilities are most important for comprehension of text, such that it would be expected for improvements in reading comprehension to be driven by increases in comprehension (listening or verbal). Hence, such findings seem consistent with the notion that intelligence provides raw material underlying the acquisition of reading comprehension skills once basic word reading skills are mastered.

Third, our results support the notion that prior Gc is important for future reading comprehension, but reading comprehension may not be essential for the accumulation of Gc. Bruck (1990) found reading comprehension differences among adults with dyslexia, despite equally poor word recognition skills. Individuals who were considered “good” comprehenders were those who had higher measured IQ as children and higher vocabulary scores as adults (also see Ferrer et al., 2010). That is, despite word recognition problems, low initial levels of reading comprehension may not necessarily either accelerate or hinder changes in intelligence or Gc, but higher intelligence or Gc may help those with poor word recognition skills and early reading comprehension deficits develop adequate reading comprehension skills. For example, children with reading disabilities who have intact language comprehension have been shown to use contextual facilitation, even more so than typical readers, to assist in reading comprehension (Nation & Snowling, 1997). Gc may facilitate the use of contextual clues, even with concurrent basic word reading deficits or fluency difficulties, although it may take time to manifest in reading comprehension skills and scores.[6]

In addition to the findings related to the dynamic parameters, there were a few interesting findings related to the covariates. For example, relative reading volume uniquely explained 3rd-grade Gc and reading comprehension scores and internal growth in Gc. Moreover, along with basic reading skills, reading volume mediated some of the influences of SES and early childhood intelligence. Although it may be argued that the measure of reading volume was just another indicator of Gc because it required accurate recall of knowledge, our findings are intriguing because reading volume is generally considered a manipulable variable (see Cunningham & Stanovich, 2001). To develop a better understanding of the influence of reading volume, researchers should consider modeling basic reading, reading volume, and vocabulary longitudinally. For example, they could test models to investigate whether higher prior basic word reading level produces gains in reading volume and if higher prior reading volume produces even greater positive changes in vocabulary. They could also test whether these potential longitudinal relations are bidirectional and dynamic. These questions are often asked by those studying a Matthew effect (Stanovich, 1986).

Last, despite the sometimes popular lore that measures of intelligence at a young age are not useful or predictive of future learning, some interesting findings emerged that were related to early childhood intelligence. To be specific, a measure of intelligence administered at approximately 24 months of age had strong total effects on both reading comprehension and Gc in 3rd grade, as well as internal growth in Gc, with the effects of sex and SES held constant. Some of these effects were mediated by reading volume and basic word reading, but even when these mediators were included, unique effects remained. In addition, a statistically significant total effect on constant reading comprehension growth was mediated by basic reading and reading volume. Therefore, those with higher intelligence scores at two years of age also had higher levels of Gc and reading comprehension in 3rd grade and had steeper constant growth in Gc and reading comprehension. Although we did not have multiple measures of fluid intelligence over time to investigate Cattell's dynamic investment hypotheses explicitly (Cattell, 1987; see Ferrer & McArdle, 2004), one simple prediction from investment theory is that, all things equal, children with higher levels of fluid intelligence, which emerges about the age of two or three, should demonstrate faster rates of learning. This simple prediction was supported in this study.

### *4.3. Limitations*

Previous research has found stronger couplings in grades 1 to 3 than in later grades (Ferrer et al., 2007). Ideally, we would have had access to information dating back to 1st grade; however, we were limited in the number of assessment time points. This limitation is likely more important for basic reading skills than for comprehension, however, because prior to 3rd grade, children are generally still in the process of learning basic reading skills.

There were a couple limitations regarding the use of the Title Recognition Task as a measure of reading volume. First, it was measured in 5th grade, so it was assumed to be a retrospective indicator, and some research supports this interpretation (Cipielewski & Stanovich, 1992). The second concern was that it may be another indicator of Gc because it required the accurate recall of information. If children were exposed equally to the books, it would be expected that those with higher Gc would be able to more accurately recall book titles. Future research should take this limitation into account. Moreover, academic motivation

[6] We should note, however, that comprehension of shorter, untimed passages (like those employed by the WJ-R Passage comprehension test), may not accurately reflect the frustrations and functional reading impairments that persist for those with specific reading disabilities.

and interest should be considered as potentially important common causes of the Gc and reading volume relations.

An additional limitation was related to the sampling of measures. Our sample of measures was limited in that we used single indicator variables. The single indicator measures were narrow in focus. Although vocabulary tests are often strong indicators of Gc (Horn & McArdle, 2007) or intelligence in general (Jensen, 1980), and Gc subtests, especially vocabulary, carry more common factor variance and less specific variance than most other intelligence subtests (Bickley et al., 1995), it would have been ideal to have multiple indicators at each time point so that composites or multiple-indicator latent variables could have been used. These limitations extend to the measure of reading comprehension, which was a cloze procedure. Regardless, construct generality is an important consideration when interpreting the findings from this research.

Last, it has been noted that the exclusion criteria for the SECCYD limits the number of children with various risk factors, including those whose first language is not English, those who were ill at birth, and those who have lived in exceedingly violent neighborhoods (e.g. NICHD ECCRN, 1996, 2004). Additionally, the ever-increasing ethnic diversity within the United States is not adequately represented, as the sample was not designed to be nationally representative. The discrepancies in representativeness are particularly observed in the White and Hispanic subgroups. Current census data indicate that over 16% of the U.S. population identify as Hispanic or Latino, compared to 6% of the study sample; while only 72.4% identify as White, versus 81.3% of the study sample (U.S. Census Bureau, 2011). Generalizations to the population, as with any sample, should be made cautiously.

### *4.4. Implications*

On the surface, one implication from our findings might be to train students to increase their vocabulary knowledge because higher Gc leads to positive change in reading comprehension. This proposition, although enticing at face value, is not necessarily supported by this study. First, we hypothesized that the Gc measure captured something broader, something akin to verbal comprehension and reasoning (Jensen, 1980). Similarly, measures of Gc typically capture breadth rather than depth of verbal knowledge (Horn & McArdle, 2007), and it is unlikely that such breadth would be systematically influenced by direct vocabulary instruction. Nevertheless, we are not speculating that building vocabulary knowledge via direct instruction is unimportant. It seems reasonable that explicit instruction with regard to domain specific vocabulary would improve comprehension. For example, reading comprehension in a specific academic domain may be improved by studying domain specific vocabulary (e.g., oncology).

The current findings, however, may have implications for the evaluation and treatment of specific reading disabilities. Gc and reading comprehension are not the same constructs. They follow different developmental patterns, with prior Gc driving positive change in reading comprehension. The development of Gc may be understood without knowing a person's reading comprehension, but development in reading comprehension is not well understood without an estimate of a person's Gc. Such a consideration is important in an evaluation of reading comprehension difficulties.

Another potentially related implication is that those with average or above average levels of Gc and who demonstrate early reading comprehension deficits may learn to compensate so that over time their reading comprehension improves on short passages. If this is true, then the type of reading interventions may need to differ between those students with a specific reading disability (e.g., average to above average Gc or intelligence with a basic word reading deficit) and those with concomitant reading comprehension (or general achievement) and intellectual deficits. For example, the two groups may respond similarly to narrowly focused word reading instruction (i.e., direct instruction of specific skills) but differently to longer-term interventions and habilitative strategies. Those with average to above average Gc may have a developed schema set or ability to comprehend and organize information in general, and thus they are more likely to develop skill at comprehending shorter reading passages despite inefficient word reading. This group, who would meet the traditional definition of a specific reading disability, may also need to focus on how to use their relative Gc or intellectual strengths to work around a basic word recognition deficit. When considering their long-term education and motivation, they may also benefit more from technology that provides them access to rigorous curricular content, consistent with their Gc or other intellectual abilities, in formats other than lengthy texts. A group with both low intellectual and general academic achievement, including a weakness in reading comprehension, may have more difficulties with organizing and generalizing knowledge in general, regardless of whether they are expected to extract it from print. Children or adolescents who demonstrate these more generalized weaknesses may need more direct instruction in basic reading skills and functional literacy, benefit more from learning strategies that help them with organizing and generalizing knowledge in general, and be motivated more by learning from texts that have immediate relevancy to them (see Shaw, 2008).

Last, future research should also investigate how interventions are appropriately evaluated. For example, if prior Gc is important in understanding changes in reading comprehension, then reading comprehension growth predictions (or slopes) will differ depending on the initial value of reading comprehension and a person's prior Gc when monitoring progress (also see Tran, Sanchez, Arellano, & Swanson, 2010). Moreover, attempting to measure positive change in reading comprehension by measures of reading comprehension only, or indirectly with measures of words read per minute, is oversimplifying the complex process observed here. Important individual differences in other intellectual or cognitive variables should be considered when monitoring the long-term progress of students.

### *4.5. Summary*

Intelligence and academic achievement are firmly linked, but the developmental dynamics between the two are less well known (Ferrer & McArdle, 2004). The developmental dynamics of Gc and reading comprehension were investigated in children and adolescents in grades 3 through 9. A unidirectional dynamic link was observed as prior Gc led to positive change in reading comprehension. Early childhood intelligence, SES, sex, basic sight word recognition, and relative reading volume had some important influences on 3rd-grade reading comprehension and Gc scores, and on internal Gc and reading comprehension growth, but they did not alter the dynamic relations. Gc is a leading indicator of positive change in reading comprehension. Gc should be considered when designing long-term reading comprehension interventions.

## References


Allen, L., Cipielewski, J., & Stanovich, K. E. (1992). Multiple indicators of children's reading habits and attitudes: Construct validity and cognitive correlates. *Journal of Educational Psychology*, *84*, 489–503.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching* (pp. 71–117). Newark, NJ: International Reading Association.

Bayley, N. (1993). *Bayley Scales of Infant Development: Second edition (standardization version).* New York, NY: The Psychological Corporation.

Benson, N. (2008). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, *26*, 27–41.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Betjemann, R. S., Wilcutt, E. G., Olson, R. K., Keenan, J. M., DeFries, J. C., & Wadsworth, S. J. (2008). Word reading and reading comprehension: Stability, overlap and independence. *Reading and Writing*, *21*, 539–558.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, *20*, 309–328.

Blaga, O. M., Shaddy, D. J., Anderson, C. J., Kannass, K. N., Little, T. D., & Colombo, J. (2009). Structure and continuity of intellectual development in early childhood. *Intelligence*, *37*, 106–113.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach.* Hoboken, NJ: Wiley.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Bradley-Johnson, S. (2001). Cognitive assessment for the youngest children: A critical review of tests. *Journal of Psychoeducational Assessment*, *19*, 19–44.

Bruck, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*, *24*, 439–454.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31–42.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

Cattell, R. B. (1987). *Intelligence: Its structure growth, and action.* Amsterdam, The Netherlands: Elsevier.

Catts, H. W., Adolf, S. M., & Ellis Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, *49*, 278–293.

Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, *54*, 74–89.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*, 264–274.

Cunningham, A. E., & Stanovich, K. E. (2001). What reading does for the mind. *Journal of Direct Instruction*, *2*, 137–149.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Deary, I. J., Whalley, L. J., & Starr, J. M. (2009). *A lifetime of Intelligence: Follow-up studies of the Scottish Mental Surveys of 1932 and 1947.* Washington, D.C.: American Psychological Association.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*, 246–262.

Ferrer, E., Hamagami, F., & McArdle, J. J. (2004). Modeling latent growth curves with incomplete data using different types of structural equation modeling multilevel software. *Structural Equation Modeling*, *11*, 452–483.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about congitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*, 935–952.

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*, 1460–1473.

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological Science*, *21*, 93–101.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2006). *The Achievement Test Desk Reference (ATDR) — Second edition: A guide to learning disability identification.* New York, NY: Wiley.

Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing*, *19*, 427–456.

Gagnon, S. G., & Nagle, R. J. (2000). Comparison of the revised and original versions of the Bayley Scales of Infant Development. *School Psychology International*, *21*, 293–305.

Gustafsson, J. -E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407–434.

Guttman, L. (1965). A faceted definition of intelligence. In R. Eiferman (Ed.), *Studies in psychology. Scripta Hierosolymitana [Jerusalem Scripts in Greek]*, *Vol.14*. (pp. 166–181) Jerusalem: Magnes Press.

Hebb, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, *85*, 275–292.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*, 127–160.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock–Johnson technical manual* (pp. 197–246). Allen, TX: DLM Teaching Resources.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253–270.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck, & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 205–247). Mahwah, NJ: Erlbaum.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: Free Press.

Kan, K. -J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, *39*, 292–302.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Tests of Educational Achievement — Comprehensive Form* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence*, *40*, 123–138.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly*, *14*, 239–262.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review*, *35*, 108–127.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*, 635–650.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Mather, N., & Woodcock, R. W. (2001). *Examiner's manual.* Woodcock–Johnson III Tests of Cognitive Abilities. Rolling Meadows, IL: Riverside Publishing.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*, 115–142.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. Collins, & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.

McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamic hypotheses of Gf–Gc theory using longitudinal life-span data. *Learning and Individual Differences*, *12*, 53–79.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical manual. *Woodcock–Johnson — III normative update*. Rolling Meadows, IL: Riverside Publishing.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, *47*, 651–675, http://dx.doi.org/10.1002/pits

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock–Johnson technical manual.* Allen, TX: DLM Teaching Resources.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*, 215–237.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide: Sixth edition.* Los Angeles, CA: Muthén & Muthén.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, *72*, 549–560.

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, *67*, 359–370.
National Institute of Child Health and Development (n.d.). NICHD Study of Early Child Care and Youth Development. Retrieved from http://www.nichd.nih.gov/research/supported/seccyd.cfm
NICHD Early Child Care Research Network. (1993). *The NICHD Study of Early Child Care: A comprehensive longitudinal study of young children's lives.* (ERIC Document Reproduction Service No. ED 353 0870).
NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, *11*, 269–306.
NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *SRCD Monographs*, *69*(4, 278), 1–146.
Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review*, *112*, 43–50.
Rohde, T. E., & Thompson, L. A. (2007). Predicting achievement with cognitive ability. *Intelligence*, *35*, 83–92.
Rupley, W. H., Willson, V. L., & Nichols, W. D. (1998). Exploration of the developmental components contributing to elementary school children's reading comprehension. *Scientific Studies of Reading*, *2*, 143–158.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
Schneider, J. W., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). (3rd ed.). New York, NY: Guilford Press.
Schwarz, G. (1978). Estimating the dimensions of a model. *The Annals of Statistics*, *6*, 461–464.
Shaw, S. R. (2008). An educational programming framework for a subset of students with diverse learning needs: Borderline intellectual functioning. *Intervention in School and Clinic*, *43*, 291–299.
Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gotay, N., et al. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*, 676–679, http://dx.doi.org/10.1038/nature04513
Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls. *Journal of the American Medical Association*, *264*, 998–1002.
Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York, NY: Macmillan.
Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.
Steiger, J. H., & Lind, J. (June). Statistically based tests for the number of common factors. *Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA*.
Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*, 383–401.
Tran, L., Sanchez, T., Arellano, B., & Swanson, L. L. (2010). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*, *44*, 283–295.
U.S. Census Bureau. (2011). State and county quick facts. Retrieved from http://quickfacts.census.gov/qfd/states/00000.html
Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407–423.
Watkins, M. W., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*, 59–68.
Wechsler, D. (1974). *Wechsler intelligence scale for children-Revised edition.* San Antonio, TX: Psych Corp.
Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence — Revised.* San Antonio, TX: Psych Corp.
Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, *8*, 231–258.
Woodcock, R. W. (2002). New looks in the assessment of cognitive ability. *Peabody Journal of Education*, *77*, 6–22.
Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.
Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery — Revised.* Rolling Meadows, IL: Riverside Publishing.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock–Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.
Yuan, K., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methods*, *30*, 165–200.