Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ _______________

Katherine Lee Date

Cognitive, school, and community factors that influence knowledge acquisition

By

Katherine Lee
Master of Public Health

Epidemiology

Timothy L. Lash, DSc
Committee Chair

Patricia Bauer, PhD
Committee Member

Cognitive, school, and community factors that influence knowledge acquisition

By

Katherine Lee

B.S., Psychology and Statistics

University of Georgia

2017

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2023

Abstract

Cognitive, school, and community factors that influence knowledge acquisition

By Katherine Lee

An important aspect of health and development is learning and building a knowledge base. People acquire new knowledge through direct experiences like reading a book or watching a documentary, or through indirect experiences like self-derivation through integration. Self-derivation is the process of integrating or combining two separate learning episodes to acquire information that was not directly taught. In this study, we investigate factors that might influence performance on this important process so that we can better understand how children learn. Based on prior research, we posit that cognitive factors like verbal comprehension, school quality factors like reading proficiency, and community factors like participation in extracurricular activities might influence self-derivation performance. We met with and collected information on 162 children between the ages of 8 and 12 years. Based on results from an exploratory factor analysis (EFA), we were able to define two latent constructs: *school* which is made up of the indicator variables math proficiency, reading proficiency, and economic disadvantage; and *cognitive* which is made up of the indicator variables verbal comprehension, visualization, and visual-auditory learning. Next, we evaluated two structural equation models (SEM) and determined that the model design based on results from the EFA was the better fitting model. Finally, we used the better fitting SEM model to predict self-derivation through integration performance. We found that the only predictor of self-derivation performance was the cognitive latent construct. This implies that the individual cognitive strategies employed by the learner are more predictive of indirect learning, as measured through self-derivation performance, above and beyond environmental factors like school quality.

Cognitive, school, and community factors that influence knowledge acquisition

By

Katherine Lee

B.S., Psychology and Statistics

University of Georgia

2017

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2023

Acknowledgements

I would like to extend a warm thank you to Dr. Patricia Bauer for giving me the opportunity to pursue this project and for providing me with the guidance, encouragement, and support needed to execute this work. I am extremely grateful for the time I have spent in the Bauer Memory at Emory Lab and for everything I have learned over the years. Thank you for almost 6 years of invaluable growth, both professionally and personally, and for molding me into the person and scientist I have become.

I would also like to thank Dr. Timothy Lash for serving as my thesis committee chair. I appreciate the time you've spent reading this work and providing thoughtful and constructive feedback on multiple drafts.

To the past and present members of the Bauer Lab- I am eternally grateful for your mentorship and emotional support. Dr. Hilary Miller-Goldwater, thank you for always taking the time to answer my questions, for providing guidance during all aspects of every project, and for your wealth of statistical knowledge. Dr. Julia Wilson, thank you for being an amazing role model- I look up to your work ethic and dedication. (Soon to be Dr.) Lucy Cronin-Golomb, thank you for learning new statistical techniques with me and for always being there to listen. Melanie Hanft Koslosky, I cannot thank you enough for countless hours of emotional support and for your invaluable contributions to this project during the design process and data collection. Alissa Miller, thank you for many, many hours of data collection and your attention to detail. Greer Spradling, thank you for your moral and emotional support, being my sounding board, and for your efforts in data cleaning and reduction.

I would also like to take a moment to acknowledge and thank my family, friends, and partner. I appreciate your love, flexibility, and kindness. I would not have been able to do this without all of you as my safety net. Your encouragement is what kept me going during the highs and lows of pursuing a master's degree and for that I am forever grateful.

Table of Contents

Introduction……………..……………………………………………………………….…1
Knowledge acquisition as measured through the self-derivation paradigm………………2
Communities and School affect learning……………………………………………….3
Current study……………………………………………………………………………4
Methods…………..………………………………………………………………………6
Participants……………………………………………………………………………...6
Stimuli and Materials…………………………………………………………………...7
Procedure………………………………………………………………………………9
Scoring…………………………………………...………………………………12
Data reduction and variable transformations…………………………………………12
Results…...……………………………………………...……………………………13
Descriptive statistics……………………………………..……………………13
Exploratory factor analysis…………………………………………………..……..14
Structural equation models predicting self-derivation…………………………...……15
Discussion……………………………………………………………………………….17
Major findings…………………………………………………………………………17
Implications………………………………………………………..………………..18
Limitations and Directions for Future Research………………………………………20
Conclusions…………………………………………………………………...….21
References………………………………………………………………………………..23
Tables and Figures…………………………………………………………………...…28
Table 1: Descriptive statistics……………………………………………………………28
Table 2: Correlation matrix……………………………………...…………………29
Table 3: Factor loadings from exploratory factor analysis …………..…………………30
Table 4: R-Squared values from each SEM model………………………………………31
Table 5: Factor loadings for Model 1……………………………………..…………..32
Table 6: Regression coefficients from SEM Model 1……………………………….……..33
Figure 1: Path diagram from exploratory factor analysis……………………….……….34
Figure 2: Scree plot for SEM Model 1………………………………………….………..35
Figure 3: Path diagram for SEM Model 1……………………………………...…..36

Cognitive, school, and community factors that influence knowledge acquisition

Introduction

Acquiring knowledge and building a knowledge base are critical to the health and well-being of an individual. To effectively navigate the world, one must successfully learn and adapt to one's surroundings. Knowledge is acquired by direct learning experiences (e.g., classrooms, watching a documentary) and by productive processes (e.g., analogy) that help a learner build their repository of knowledge. Knowledge accumulation is associated with higher educational attainment, which in turn is linked to higher socioeconomic status (SES) in adulthood (Adler et al. 1994; Anderson & Armstead 1995; Halleröd & Gustafsson 2011; Yu & Williams 1999). As well as learning outcomes like academic achievement relate to future health outcomes (Le-Scherban et al., 2014). Given that SES is predictive of physical and mental health and overall well-being, knowledge accumulation in childhood is an important aspect of current and future health (Chen, Martin, & Matthews, 2007; Zimmer, Hanson, & Smith, 2016; Smith, 2005). There is large individual variability on knowledge accumulation measures, and there are a number of factors that relate to this variability. For example, certain cognitive factors such as verbal comprehension are highly predictive of how well one performs on tests of knowledge accumulation (Varga, Esposito, & Bauer, 2019). In addition, the quality of schools/education and the environment/community in which children are raised influence academic achievement in general (Eamon, 2005; Covay & Carbonaro, 2010, Epstein, 2010). What has yet to be tested is whether this important aspect of health, namely knowledge accumulation, is influenced by the community in which children learn (i.e., involvement in extracurricular activities), by characteristics of the schools that children attend (i.e., reading proficiency, math proficiency, number of students per teacher), or both. In addition, it will be important to examine how these community and school characteristics influence knowledge accumulation in relation to the individual cognitive factors the learner employs. These are the questions addressed in the present research.

Knowledge acquisition as measured through the self-derivation paradigm

One of the most important means of knowledge accumulation is productive processes such as analogy, deduction, generalization, induction, associative and transitive inference (e.g., Dias & Harris 1988; Gentner, 1989; Goswami, 2011), and the focus of this research: *self-derivation through memory integration* (Bauer & San Souci, 2010). Productive processes are important because the learner goes beyond what is directly taught to produce new knowledge. Self-derivation through memory integration is particularly important because this paradigm uses naturalistic stimuli such as sentences, passages, diagrams, and other means of acquiring knowledge that are more applicable to the way people learn outside of the laboratory. For example, during a learning opportunity (perhaps while reading a pamphlet) one might learn the fact that apple seeds are called pips, then during another separate learning opportunity (perhaps in a chemistry class) one may learn that pips contain cyanide. When challenged with the question -what do apple seeds contain? – the answer can be self-derived by *integrating* the two learning opportunities: cyanide. The learner was never directly taught the fact that apple seeds contain cyanide, but this new self-derived fact represents the knowledge base expansion/learning that happens during productive processing.

As people learn, grow, and develop, we participate in self-derivation through integration, and much research has been done to determine the implications it has on different characteristics of learning. We know that children as young as 4 years old are successful at self-derivation (Bauer & San Souci, 2010). In addition, we know that self-derivation occurs across a range of sematic topic areas including but not limited to arts, humanities, sciences, history, and information about prescription medications (Bauer et al., 2019; Dugan & Bauer, 2022; Esposito et al., 2021; reviewed in Bauer, 2021). Information that is self-derived is quickly incorporated into the knowledge base (Bauer & Jackson, 2015), retained over time (Varga et al., 2019b; Varga & Bauer, 2013), can itself be used productively (Wilson & Bauer, 2021), and occurs under naturalistic educational experiences such as virtual museum exhibits (Cronin-Golomb & Bauer, 2022). Research on self-derivation has also been conducted in classrooms, which has revealed that self-derivation performance is related to academic achievement (Esposito & Bauer, 2017, 2022) and predicts end-of-year reading comprehension and math achievement scores for children in Grades 1-3 (Esposito & Bauer, 2017). We also know there are several cognitive factors that relate to self-derivation performance such as verbal comprehension and working memory (Varga, Esposito, & Bauer, 2019; Bauer et. al, 2023 [submitted]).

Communities and School affect learning

It is intuitive that the environment in which one is raised influences learning, academic achievement, and therefore the overall health of an individual. Two major components of a child's environment are the school they attend and the larger community where they participate in activities. School and community characteristics have each been measured in myriad ways. Yet, overwhelmingly, there is a consensus that the characteristics of both the school and community influence students' learning and overall academic achievement.

A number of studies investigate the effect of school quality and/or characteristics on student academic achievement. Catsambis & Beveridge (2012) found that even after controlling for individual background variables, disadvantaged schools are directly related to lower levels of mathematics achievement in a sample of eighth grade students. Koc & Celik (2014) found a moderate negative correlation between student-teacher ratio and student achievement, meaning that cities with more students per teacher tend to have lower student achievement levels. Marshall (1993) discusses that even if there are great individual teachers within a school, it is also important that the school operates as a coherent unit to ensure the success of the students. Marshall (1993) argues that good teaching is not enough, and in order to make a difference there must be a school-wide effort, and this cohesive effort usually starts with the principal and instructional leadership.

In addition to school quality, there are many studies that examine the relationship between characteristics of the community and student academic achievement. Eamon (2005) found that residence in higher quality neighborhoods was related to higher reading achievement in a sample of adolescents. Covay and Carbonaro (2010) explored the relationship between SES and academic achievement in elementary students by investigating the role of extracurricular participation. They found that students from higher SES backgrounds participate more in extracurricular activities, more participation in extracurricular activities improves student achievement, and the SES advantage on academic achievement is moderated by participation in extracurricular activities.

Current study

To see how cognitive abilities, school characteristics, and community characteristics relate to self-derivation performance, we collected data on 162 children ages 8 to 12 years. In addition to testing self-derivation through memory integration, we administered tests of cognitive abilities using the Woodcock-Johnson IV III and Woodcock-Johnson IV Tests of Cognitive Abilities (Schrank et al., 2014). Specifically, we tested verbal comprehension (Woodcock-Johnson IV III, Tests 1A-1D), visual-auditory learning (Woodcock-Johnson IV, Test 13), and visualizations (Woodcock-Johnson IV, Tests 7a and 7b). In addition, we administered a survey filled out by the caregiver to collect information on the school the child attends, the child's involvement in extracurricular activities, and caregiver(s) education/career information.

Given the large number of measures, we first performed an exploratory factor analysis to determine whether any of the variables factor together into latent constructs. We anticipated that there would be three latent variables: cognitive, school, and community. We expected the *cognitive* latent variable to be comprised of verbal comprehension, visual auditory learning, and visualizations; the *school* latent variable to be comprised of math proficiency, reading proficiency, and students per teacher; and the *community* latent variable to be comprised of number of extracurricular activities, economically disadvantaged, and SES. After defining latent variables, we compared two structural equation models to determine the one that best fits our data, and then used that model to predict self-derivation through integration performance. Based on prior research, we expected that the cognitive latent variable would predict self-derivation performance and this latent variable would interact with school and community variables to predict self-derivation performance.

We selected the age range of 8 to 12 years old for several reasons. First, to date, children younger than 8 have only been tested on the self-derivation through memory integration task using a story passage paradigm (facts are presented as part of a story to keep children engaged) and with a lower number of trials (typically 4). Children 8 years or older can and have been tested on the self-derivation task using a single sentence paradigm (the fact is displayed and/or read out loud for them), thus taking less time and permitting more trials per session (typically 8-20). This age range also permitted the opportunity to look at school-level variables from grades 2 to 8.

In summary, in the present research we collected self-derivation through memory integration performance, multiple cognitive measures, and information on the school and community in which 8- to 12-year-old children live. We expect to define three latent variables in this data set and based on prior research, we anticipate the cognitive latent variable will predict self-derivation performance and indicators of higher quality schools (e.g. higher reading and math proficiency) and communities (e.g. more involvement in extracurricular activities) will predict and/or interact with the cognitive latent variable to predict self-derivation performance.

Method

Participants

Participants were 162 children ages 8 to 12 years (*M* = 10.47, *SD* = 1.37, range = 8.13-12.91 years). Participants were recruited in a variety of ways: most were recruited through a Child Study Center database of families who had previously expressed interest in participating in research studies, a third-party marketing firm, or by referral of other participants enrolled in the study. The sample was 86 female (53%) and 74 male (47%); 2 caregivers did not report their children's gender. Based on caregiver report, the sample was 6% Asian, 15% Black, 1% Middle Eastern or Arab, 66% White, 5% mixed race, and 7% did not report on their children's race. Nine percent of the sample self-identified as Hispanic or Latinx. Ninety percent of caregivers had received at least some college education, and 59% of caregivers had received at least some graduate level education; 4% did not report on caregiver education. An additional 11 children were recruited into the study, but their data were not included because of failure to complete the second test session (3), technical failure (3), prior participation in a related study (1), caregiver report of a developmental disability (2), and child-initiated request to end the session before all tasks were administered (2).

Participants were recruited to take part in a larger study, which consisted of two online test sessions in each of two consecutive years (2 sessions in Wave 1 and 2 sessions in Wave 2). Data collection was initiated during the 2020 COVID-19 related shutdown, thus necessitating an online protocol. The present study only uses data collected in Wave 1. Sessions took place an average of 7 days apart (range = 4-9 days). Before the beginning of the first test session, caregivers provided written informed consent for their children to participate. Children provided verbal (ages 8-9) or written (ages 10-12) assent at the beginning of the first session. Participants were compensated with $40.00 in an e-gift card at the end of the second session. The procedures were reviewed and approved by the Emory University Institutional Review Board.

Stimuli and Materials

The full protocol included tests of self-derivation through memory integration, retention of self-derived facts, retention of explicitly taught facts, measures of achievement, and measures of cognitive abilities. The focus of this report is on self-derivation, measures of cognitive abilities, and results from the survey filled out by the caregiver (school and community characteristics). Description and analysis of measures of achievement and retention of self-derived facts are beyond the scope of this report. For present purposes, they served as buffer activities, as described below.

Self-derivation through memory integration. The stimuli were 40 pairs of related facts that could be used to self-derive new facts (hereafter, *self-derived facts*). All of the stimuli were true facts. The facts were pilot tested with adults to ensure they were unfamiliar. It is reasonable to assume that if the facts were unfamiliar to adults, they also would be unfamiliar to children, and thus that children would be experiencing the facts for the first time in the context of the test session. Pilot testing also ensured that the production of the self-derived facts was dependent on experiencing both members of the fact pairs, and thus that their production was the result of self-derivation through integration. The stimuli are available through the Bauer Lab Integration and Self-derivation Stimulus (BLISS) bank (Bauer, 2020: BLISS bank stimulus numbers S002, S055-057, S066-67, S069, S084, S086, S093, S096-97, S108-111, S113, S126-147). In addition to the individual and self-derived facts, stimuli also included 24 "filler" facts (BLISS bank stimulus numbers F098-102, F104, F106-107, F109-110, F119-132). Filler facts could not be integrated with one another to derive new facts and were used as the source of foils for forced-choice testing.

Measures of cognitive abilities. We administered multiple tests of cognitive abilities using the Woodcock-Johnson IV III and Woodcock-Johnson IV Tests of Cognitive Abilities. Specifically, we tested verbal comprehension (Woodcock-Johnson IV III, Tests 1A-1D), visual-auditory learning (Woodcock-Johnson IV, Test 13), and visualizations (Woodcock-Johnson IV, Tests 7a and 7b). To accommodate online data collection, the tests were rendered as Qualtrics® surveys.

Caregiver survey. We asked the caregiver to fill out an online survey on the RedCap platform (IRB approved to collect identifiable information) during the week between meetings with the participant. For the purposes of this research we focus on the following questions from the caregiver survey: child date of birth, child gender identity, name of the school the child attends, grade level of child, caregiver 1 education level (with menu options: some high school, high school, some college, technical or AA degree, college degree, some graduate, post graduate), caregiver 1 area of occupation (with menu options: student; homemaker; office, food service, or retail staff; skilled trade/technical; professional/managerial; self-employed; nursing/health services; unemployed, retired, or other), caregiver 2 (if applicable) education level, caregiver 2 (if applicable) area of occupation, whether the child had music training and if so, how many years, a list of 25 (select all that apply) extracurricular activities/clubs, and whether the child plays any sports. Using the name of the school the child attends, we were able to obtain proxy measures of math proficiency, reading proficiency, economical advantage, and students per teacher, (information found: https://www.usnews.com/education/k12/elementary-schools/georgia, accessed in 2022).

Procedure

Participants took part in two online sessions separated by approximately 1 week (*M* = 7.13; range = 4-9 days). Sessions were conducted online, via Zoom, and were recorded. Children were tested by one of six experimenters and each child had the same experimenter at both sessions. The experimenters followed a detailed written protocol, fidelity to which was assessed by regular group viewing of the session recordings. The protocols at the two sessions are as follows.

Session 1

Self-derivation through integration. Children were presented with 42 facts, presented across two learning phases. Twenty facts (10 fact pairs) were presented in a condition where the participants learned both related facts in a pair (Integration condition). These facts were related in pairs such that when the two facts in a pair were integrated with one another, they could be used to answer an integration question by self-deriving a new fact that had not been explicitly presented. Ten additional facts were presented in a 1-fact control condition where the participants only learned one of the facts in a given pair. The remaining 12 facts were fillers whose purpose was to provide alternatives for forced-choice testing and an independent assessment of fact recall. The current work does not analyze facts in the control condition or filler facts. For information about these control and filler facts see Bauer et al. (2023, submitted).

In Learning phase 1, children were presented with 21 facts including the first member of 10 pairs of facts that could be integrated (the other 11 facts were facts in the control condition or filler facts, not analyzed in the current work). Children were instructed to pay attention to the facts because they might be asked some questions about them later. They were not advised that any of the facts were related to one another or that some of the facts could be combined to derive new facts. Each fact was displayed one at a time as a sentence on the Zoom screen via a Qualtrics survey. While the fact was displayed, the experimenter read it aloud and then the child repeated the fact. After all 21 facts had been displayed and read aloud, children participated in the test of verbal comprehension (Woodcock-Johnson III tests 1A-1D) followed by learning phase 2. The second learning phase featured display and reading of the second members of the 10 pairs of facts in the integration condition (and additional facts not analyzed here). Again, children were not advised that any of the facts were related to one another or that some of the facts could be combined to derive new facts. After all 21 facts had been displayed and read aloud, children participated in a buffer activity requiring approximately 10 minutes.

After the buffer activity, children were tested for self-derivation through integration in both open-ended and forced-choice formats (they also were tested for open-ended recall of 15 of the control facts and 10 of the filler facts, not analyzed here). They first were asked 20 open-ended integration questions to test for self-derivation: 10 for fact pairs presented in the integration condition and 10 for facts from the control condition. At no time were children advised that they could answer the questions by forming relations between the facts presented in the learning phases. After open-ended testing for self-derivation, children were tested for open-ended recall of the control and filler facts. Open-ended testing was followed by forced-choice testing of integration questions that were not answered correctly in open-ended testing.

In total, there were 40 pairs of related stem facts, divided into two sets of 20 pairs of related facts. Half of the participants were tested on one set of stem facts and half on the other. Each set of facts was used approximately equally often across participants. Within a set, each fact pair was used in the integration and control conditions approximately equally often, and in the integration condition, each member of the fact pair was presented in Learning phase 1 and Learning phase 2 approximately equally often. Each fact pair was presented in one of four different random orders, each used approximately equally often across participants. Participants were pseudo-randomly assigned to one of the four orders, constrained by the need to use each order approximately equally often. The order of presentation of open-ended integration questions was randomized in Qualtrics; the order of forced-choice testing of integration questions not answered correctly in open-ended testing also was randomized in Qualtrics. Participants were also asked open-ended questions for recall of the control and filler facts.

After testing self-derivation through integration and fact recall, children participated in a 10-minute buffer activity. Immediately after the buffer activity, children were administered Woodcock-Johnson IV Test 7a.

Session 2

Session 2 took place roughly 1 week after Session 1. Children were tested for retention of the self-derived facts from Session 1 (not analyzed in the current work), followed by tests of cognitive abilities (Woodcock-Johnson IV Test 7b and Woodcock-Johnson IV, Test 13) and additional buffer activities (not analyzed in the current work).

Scoring

For self-derivation through integration, one point was awarded for each correct response, for a total possible of 10 self-derived facts and a proportion correct score was calculated. The WCJ-III and WCJ-IV cognitive abilities tests (verbal comprehension, visual-auditory learning, and visualizations) were scored per the test protocol and standardized using WCJ proprietary software. SES was calculated using a combination of caregivers' occupation (0 = unknown or unemployed; 1= homemaker, student, retired, other; 2 = office, food service, or retail staff; 3 = skilled trade/technical; 4= nursing/health services; 5 = self-employed; 6 = professional/Managerial) and education (0= unknown or none; 1 = some high school; 2 = high school; 3 = some college; 4 = technical or AA degree; 5 = college degree; 6 = some Graduate; 7 = post graduate) numbers assigned to each level were summed together for a total SES score.

Data reduction and variable transformations

Due to the nature of the planned analysis, we first needed to transform variables that were "reverse coded" to meet the assumption that the indicator variables for a given latent variable are positively correlated. In addition, indicator variables should be on the same scale, therefore variables that were originally proportions or decimals were transformed to integers. Transformed variables were used in all results and analyses.

The following variables were transformed:

Economically disadvantaged: Economically disadvantaged is a variable obtained through: https://www.usnews.com/education/k12/elementary-schools/georgia (accessed in 2022) which is a measure of the percentage of economically disadvantaged students at a given school. Given that the percentage of economically disadvantaged students negatively correlated with other indicator variables predicted to factor into the school latent variable, we reversed economically disadvantaged by subtracting the percentage from 100.

Students per teacher: Students per teacher is a variable obtained through: https://www.usnews.com/education/k12/elementary-schools/georgia (accessed in 2022) and is the average number of students per teacher at a given school. Given that this variable negatively correlated with other indicator variables predicted to factor into the school latent variable, we reversed this variable. We reversed the number of students per teacher by taking 1+maximum (1+29 = 30) and subtracting the result from 30.

Results

Results are presented in three sections. First, we provide descriptive and summary statistics for the variables of interest. Next, we report results from an exploratory factor analysis to determine and define latent variables. Finally we present results from structural equation models predicting self-derivation through memory integration performance. All analyses were conducted using JASP software.

Descriptive statistics

Descriptive statistics are provided in Table 1. We determined that each of the variables met the assumptions needed to assume a normal distribution based on prior established criteria of skewness (meets criteria if between -2 and 2) and kurtosis (meets criteria id between -7 and 7). Missing data are attributed to either lack of information found about the school online, questions left blank on the caregiver survey, and/or procedural error during data collection.

Exploratory factor analysis

We conducted an exploratory factor analysis using the minimum residual estimation method and the following settings in JASP: parallel analysis of factors, oblique rotation to account for high correlations between variables, analysis based on the correlation matrix, factor loading cut off set at 0.4, and handling missing values by excluding cases pairwise. The following variables were included in the exploratory factor analysis: verbal comprehension, visual auditory learning, visualization, math proficiency, reading proficiency, economically disadvantaged, students per teacher, number of extracurricular activities, and SES. First, we checked that all assumptions were met to conduct an exploratory factor analysis: all variables included are continuous, the data follow a normal distribution (see above for skew and kurtosis), and there is a linear relationship between the variables and the factors as determined by a Kaiser-Meyer-Olkin overall MSA value = 0.692 (meets criteria if above 0.5) and a Bartlett's test χ^2 = (36) = 607.02, p-value < 0.001 (meets criteria if p-value is significant at α=0.05). In addition, variables predicted to factor together are positively correlated (Table 2).

We first tested the fit of the model by conducting a chi-squared test. In this analysis, model fit is rejected (bad fit) when the chi-squared test is significant at α=0.05. The results from our model fit test indicated that we failed to reject our model fit χ^2 = (19) = 30.18, p-value = 0.050. However, because the chi-squared test approached significance, we conducted additional measures of fit by using the additional fit indices provided by JASP. We found a root mean square error of approximation (RMSEA) = 0.06, 95% confidence interval = (0.003, 0.099), and following guidelines provided by Browne and Cudeck (1993), we determined this model has an acceptable fit (values less than 0.08 are acceptable, values greater than 0.1 should be rejected).

Results of the exploratory factor analysis suggest that there are two latent variables in the dataset. The first latent variable includes the indicator variables of reading proficiency, math proficiency, and economically disadvantaged. The second latent variable includes the indicator variables of verbal comprehension, visual-auditory learning, and visualization. SES, students per teacher, and number of extracurricular activities did not factor with any other variables (factor loadings can be found in Table 3). Visualization of the path diagram can be found in Figure 1, where wider arrows indicate stronger factor loading onto the corresponding latent variable. Inspection of the Scree plot (Figure 2) further indicates that there are two latent variables in this dataset. Based on these results, we have defined the first latent variable as a measure of school quality (school) and the second latent variable as a measure of cognitive ability (cognitive) of the participant.

Structural equation models predicting self-derivation

We conducted two structural equation models (SEM) to predict self-derivation through memory integration performance. Model 1 is designed while taking the results from the exploratory factor analysis (above) into account with two latent variables: s*chool* with math proficiency, reading proficiency, and economically disadvantaged; and *cognitive* including verbal comprehension, visual auditory learning, and visualizations as indicators. We also included an additional independent variable in Model 1, namely SES, due to our theoretical predictions and large body of literature supporting that SES influences learning outcomes (for meta-analysis see Sirin, 2005). Model 2 is our predicted model structure with three latent variables: *cognitive* with verbal comprehension, visual auditory learning, and visualizations as indicators; *school* with math proficiency, reading proficiency, and students per teacher as indicators; and *community* with number of extracurricular activities, economically disadvantaged, and SES as indicators. Importantly, the differences between Models 1 and 2 are that (a) Model 2 includes an addition latent variable (community) while Model 1 has two latent variables and an independent variable (SES), (b) Model 2 includes economically disadvantaged as an indicator in the community latent variable, while Model 1 includes economically disadvantaged is an indicator in the school latent variable, and (c) Model 2 includes students per teacher in the school latent variable and number of extracurricular activities in the community latent variable while Model 1 does not include these indicator variables.

To select among the models, we first evaluated measures of fit for both models to determine which one is most appropriate for our data. A chi-squared model fit test revealed that both models adequately fit the data χ^2 (18) = 21.49, p-value = 0.256 (Model 1) and χ^2 (30) = 40.33, p-value = 0.099 (Model 2) where bigger p-values indicate a larger probability that the source data fits the number of factors specified. In addition, there is no significant loss in model fit when comparing Model 2 to Model 1 χ^2 (12) = 18.84, p-value = 0.092. Comparative fit index (CFI) also revealed that both models fit the data with Model 1 CFI = 0.991 and Model 2 CFI = 0.974 (CFI above 0.90 is recommended). Lastly, we evaluated root mean square error of approximation (RMSEA) for Model 1 RMSEA = 0.035 and Model 2 RMSEA = 0.046 (recommended for RMSEA values to be below 0.04). Although our predicted model (Model 2) passes all but one of the model fit tests, we believe Model 1 is the best fit for the data given the results from the exploratory factor analysis and RMSEA values. R-squared values for both models can be found in Table 4.

Moving forward with Model 1, factor loadings for each of the indicator variables are significant for their given latent variable (Table 5). The path diagram is available in Figure 3. Results from Model 1 revealed that the cognitive latent variable significantly predicts self-derivation through memory integration performance, but the school latent variable and SES are not significant predictors of self-derivation performance (regression coefficients and p-values available in Table 6). These results suggest that cognitive factors as measured through verbal comprehension, visualizations, and visual-auditory learning indicator variables are predictive of self-derivation performance above and beyond school factors and SES. This suggests that the individual cognitive factors that the learner employs is the strongest predictor of knowledge accumulation as measured through self-derivation through memory integration.

Discussion

The major purposes of the present research were to 1) find and define latent variables that address constructs of individual cognitive abilities, school characteristics, and community characteristics, and 2) determine whether these latent constructs predict self-derivation through memory integration performance.

Major Findings

The first question addressed in the current research was to define latent constructs within the dataset. Based on prior research, we hypothesized three latent constructs of individual cognitive abilities, school characteristics, and community characteristics. We theorized that the latent construct of individual *cognitive abilities* would include verbal comprehension, visual-auditory learning, and visualizations as indicator variables; the *school* latent construct would include math proficiency, reading proficiency, and number of students per teacher; and the *community* latent construct would include number of extracurricular activities, economically disadvantaged, and SES. Based on an exploratory factor analysis, our predictions were only partially supported. As expected, we were able to define a cognitive latent construct that included verbal comprehension, visual-auditory learning, and visualizations. We were also able to define a school latent construct that included math and reading proficiency; however, number of students per teacher did not factor into the school latent construct as predicted and economically disadvantaged did. Contrary to our predictions, we were unable to define a community latent construct. In addition, number of students per teacher, SES, and the number of extracurricular activities did not factor with any other measure.

Another contribution of the current work is to establish a structural equation model that fits our dataset. We evaluated two SEM models - Model 1 was designed with the two latent constructs defined based on results from the exploratory factor analysis and also included SES, and Model 2 was designed using prior literature and our hypothesized latent constructs. Both models fit the dataset adequately, but ultimately, we determined that Model 1 was the better fitting model for our data due to model fit indices.

The second goal of the current work was to use structural equation modeling to predict knowledge acquisition as measured through self-derivation through memory integration performance. Using our best fitting model, Model 1, we found that the only significant predictor of self-derivation performance was the cognitive latent construct. The school latent construct and SES were not significant predictors of self-derivation performance. This suggests that individual cognitive factors are more highly predictive of knowledge acquisition, above and beyond measures of school and SES.

Implications

Overall, this work contributes to our current understanding of an important cognitive process by which knowledge accumulates, namely self-derivation through memory integration.

Prior research has established that there is large individual variability in self-derivation performance and that this process relates to other cognitive measures such as verbal comprehension (Bauer et al., 2023 [submitted]; Varga, Esposito, & Bauer, 2019; Varga & Bauer, 2017). This is consistent with the finding in the current work that the cognitive latent construct, which includes verbal comprehension as an indicator, predicts self-derivation performance.

Contrary to our predictions, we were not able to define a community latent construct. In prior research, SES is typically positively correlated with participation in extracurricular activities, SES is positively correlated with student achievement and learning outcomes, and greater participation in extracurricular activities is related to higher student achievement and learning outcomes. Based on these prior findings, we predicted that participation in extracurricular activities and SES would factor together into a latent construct, and that latent construct would predict our learning outcome of self-derivation. However, in the current work these variables did not factor together, and in fact were not even correlated with each other (Table 2). We believe this may be due to how participation in extracurricular actives was operationalized in the current work. Participation in extracurricular activities was a measure of the *number* of extracurricular activities reported by the caregiver, and perhaps a more sensitive measure for future studies would be the *number of hours per week* the child engages in extracurricular activities.

Due to the large body of literature suggesting SES predicts academic achievement and learning, we included SES in Model 1 even though it did not factor into a latent construct to determine its relationship with self-derivation through memory integration. Our findings are not consistent with this prior work because we did not find that SES predicted self-derivation performance. We instead interpret this as consistent with the line of work suggesting that self-derivation performance is highly individualistic and more indicative of the cognitive processes the learner employs rather than environmental factors such as SES.

In a similar vein, we did not find a relation between the school latent construct and self-derivation performance. We interpret this the same way: that environmental factors such as school quality as measured by math proficiency of the school, reading proficiency of the school, and percentage of the school who are economically disadvantaged are not as predictive of self-derivation performance because self-derivation is a more individualistic process.

We believe what makes self-derivation through memory integration different from other learning outcomes is the productive nature of the self-derivation process. Learning/accumulating knowledge through self-derivation is a *productive* process that requires the learner to go beyond what is directly taught to put together two separate pieces of information, which is different from other learning outcomes that might instead measure what is directly taught.

Limitations and Directions for Future Research

The current work is not without limitations. First, data for this study were collected via Zoom meetings with participants rather than having participants come in person to the laboratory. Zoom meetings were necessitated due to the Covid-19 pandemic, but we acknowledge that there are challenges associated with meeting with participants online rather than meeting with them in person. The laboratory is a controlled space with limited distractions, and meeting with children online did not allow us to have full control over environmental distractions such as siblings or other people in the household talking, or participants having access to toys or other possible distractions during the session. We suspect these distractions were minimal, especially given that most of these participants were used to the Zoom platform due to remote schooling that was also necessitated by the Covid-19 pandemic. Despite these challenges, one benefit of remote data collection is that we had the opportunity to meet with children who lived further from the university and who otherwise might not have had the means to commute and participate in person. Ultimately, we are not concerned with the remote nature of data collection because testing conditions were the same across all participants and any potential challenges impacted the full sample.

Another limitation of the current work is that the R^2 value for the visualizations indicator variable loading onto the cognitive latent variable is below the recommended threshold (0.4), which might suggest that visualization is not a strong indicator of the cognitive latent variable (Table 4). In future studies, it might be beneficial to replace this indicator with a different measure of cognition or perhaps include a larger battery of cognitive measures to possibly find different types of cognitive latent constructs.

Inspection of Table 6 shows that estimates are low for SEM Model 1 (low beta weight), so although the cognitive latent variable still significantly predicts self-derivation performance, this might indicate that it is not a particularly strong predictor. Future studies are needed to determine other possible predictors of self-derivation performance to find what might be contributing to the high individual variability in performance.

Conclusion

In conclusion, the current work provided the opportunity to define latent constructs in a dataset with a large number of variables to predict an important learning outcome, namely self-derivation through memory integration. We defined two latent constructs, individual cognitive abilities and school quality, which are important constructs to the overall health and wellbeing of children and highly indicative of future health. We found that the cognitive latent construct was the only significant predictor of self-derivation performance, above and beyond school quality and SES. The findings imply that self-derivation performance is likely more individualistic than other more direct measures of learning.

References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist, 49*(1), 15–24. https://doi.org/10.1037/0003-066X.49.1.15

Anderson, N. B., & Armstead, C. A. (1995). Toward understanding the Association of Socioeconomic Status and Health. *Psychosomatic Medicine*, *57*(3), 213–225. https://doi.org/10.1097/00006842-199505000-00003

Bauer, P. J. (2021). We know more than we ever learned: Processes involved in accumulation of world knowledge. *Child Development Perspectives, 15(4),* 220-227. https://doi.org/10.1111/cdep.12430

Bauer, P. J., Cronin-Golomb, L. M., Porter, B. M., Jaganjac, A., & Miller, H. E. (2021). Integration of memory content in adults and children: Developmental differences in task conditions and functional consequences. *Journal of Experimental Psychology: General, 150*(7), 1259–1278. https://doi.org/10.1037/xge0000996

Bauer, P. J., Dugan J. A., Cronin-Golomb, L. M., Lee, K. A., Del Solar, B., Hanft, M., & Miller, A. (Submitted 2023, invited revise resubmit). Development of self-derivation through memory integration and relations with world knowledge. *Child Development*

Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 271–282. https://doi.org/10.1037/a0037405

Bauer P. J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4- to 8-year-old children. *Child Development, 88*, 247-262. doi.org/10.1111/cdev.12584

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology*, 107, 452-465. (Ms. ID# NIHMS210076)

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA: Sage.

Catsambis, S., & Beveridge, A. A. (2001). Does neighborhood matter? family, neighborhood, and school influences on eighth-grade mathematics achievement. *Sociological Focus, 34*(4), 435–457. https://doi.org/10.1080/00380237.2001.10571212

Chen, E., Martin, A. D., & Matthews, K. A. (2007). Trajectories of socioeconomic status across children's lifetime predict health. *Pediatrics*, *120*(2). https://doi.org/10.1542/peds.2006-3098

Covay, E., & Carbonaro, W. (2010). After the Bell. *Sociology of Education*, *83*(1), 20–45. https://doi.org/10.1177/0038040709356565

Cronin-Golomb, L. M., & Bauer, P. J. (2022). Support for learning under naturalistic conditions. *Cognitive Research: Principles and Implications*, *7*(1). https://doi.org/10.1186/s41235-022-00435-0

Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, *6*(3), 207-221. doi: 10.1111/j.2044-835x.1988.tb01095.x

Dugan, J. A., & Bauer, P. J. (2022). Putting the pieces together: Cognitive correlates of self-derivation of new knowledge in elementary school classrooms. *Journal of Experimental Child Psychology*, 221, 105441. doi: 10.1016/j.jecp.2022.105441

Eamon, M. K. (2005). Social-demographic, school, neighborhood, and parenting influences on the academic achievement of Latino Young Adolescents. *Journal of Youth and Adolescence*, *34*(2), 163–174. https://doi.org/10.1007/s10964-005-3214-x

Epstein, J. L. (2010). School/Family/Community Partnerships: Caring for the children we share. *Phi Delta Kappan*, *92*(3), 81–96. https://doi.org/10.1177/003172171009200326

Esposito, A. G., & Bauer, P. J. (2022). Determinants of elementary-school academic achievement: Component cognitive abilities and memory integration. *Child Development, 93(6),* 1777-1792. doi: 10.1111/cdev.13819

Esposito, A. G., & Bauer, P. J. (2017). Going beyond the lesson: Self-generating new factual knowledge in the classroom. *Journal of Experimental Child Psychology*, *153,* 110-125. doi.org/10.1016/j.jecp.2016.09.003

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 199-241). New York, NY, US: Cambridge University Press. doi: 10.1017/cbo9780511529863.011

Goswami, U. (2011). Inductive and deductive reasoning. In U. Goswami (Ed.), *Childhood Cognitive Development* (pp. 399–419). Oxford, UK: Wiley–Blackwell. doi.org/10.1016/j.tics.2010.10.001

Halleröd, B., & Gustafsson, J.-E. (2011). A longitudinal analysis of the relationship between changes in socio-economic status and changes in health. *Social Science & Medicine, 72*(1), 116–123. https://doi.org/10.1016/j.socscimed.2010.09.036

JASP Team (2023). JASP (Version 0.17.3)[Computer software]

Koc, N., & Celik, B. (2015). The impact of number of students per teacher on student achievement. *Procedia - Social and Behavioral Sciences*, *177*, 65–70. https://doi.org/10.1016/j.sbspro.2015.02.335

Marshall, K. (1993). Teachers and Schools: What Makes a Difference: A Principal's Perspective. *Daedalus*, *122*(1), 209–242. http://www.jstor.org/stable/20027156

Schrank, F. A., McGrew, K.S., & Mather, N. (2014). Woodcock-Johnson IV Tests of Cognitive Abilities. Rolling Meadows, IL: Riverside.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A Meta-analytic review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Smith, J. (2005). The impact of SES on health over the life-course. *Journal of Human Resources*. https://doi.org/10.7249/wr318

Varga, N. L., & Bauer, P. J. (2013). Effects of delay on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology,* 115, 326-341. doi:10.1016/j.jecp.2013.01.008

Varga, N. L., Esposito, A. G., & Bauer, P. J. (2019b). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General, 148(4),* 739-762. doi.org/10.1037/xge0000581

Wilson, J. T., & Bauer, P. J. (2021). Prompt-facilitated learning: The development of unprompted memory integration and subsequent self-derivation. *Memory & Cognition, 49(7),* 1473-1487. doi: 10.3758/s13421-021-01155-4.

Yu, Y., Williams, D.R. (1999). Socioeconomic Status and Mental Health. In: Aneshensel, C.S., Phelan, J.C. (eds) Handbook of the Sociology of Mental Health. Handbooks of Sociology and Social Research. Springer, Boston, MA. https://doi.org/10.1007/0-387-36223-1_8

Zimmer, Z., Hanson, H. A., & Smith, K. (2016). Childhood socioeconomic status, adult socioeconomic status, and old-age health trajectories. *Demographic Research*, *34*, 285–320. https://doi.org/10.4054/demres.2016.34.10

Table 1: Descriptive statistics

	n	Missing	Mean	Std. Dev	Skewness	Kurtosis	Min	Max
SES	155	7	19.87	4.72	-0.73	0.28	6	26
Extra curr. act	155	7	2.07	1.78	1.12	1.38	0	9
Verbal Comp	157	5	107.27	10.77	0.06	-0.58	82	133
Visualization	162	0	104.21	13.25	0.01	0.48	65	139
Visual-Auditory Learning	160	2	104.84	13.51	0.02	0.28	72	150
Students per teacher	137	25	16.26	4.44	-1.30	2.96	1	26
Math Prof	99	63	61.43	17.39	-0.33	0.02	20	98
Reading Pro	99	63	63.69	15.44	-0.15	-0.40	24	100
Econ Disadv	103	59	74.18	21.85	-1.25	1.59	0	100
Self-derivation	162	0	0.33	0.21	0.41	-0.38	0	0.8

Table 2: Correlation table

Variable		SES	Extra curr. act	Verbal Comp	Vis	Vis-Aud Learning	Students per teacher	Math Prof	Reading Prof	Econ Disadv	Self derivation
SES	Pearson's r	—									
	p-value	—									
Extra curr. act	Pearson's r	0.10	—								
	p-value	0.24	—								
Verbal Comp	Pearson's r	0.21	0.15	—							
	p-value	0.01	0.08	—							
Vis	Pearson's r	0.21	0.06	0.35	—						
	p-value	0.01	0.49	< .001	—						
Vis Aud Learning	Pearson's r	0.18	0.18	0.52	0.36	—					
	p-value	0.02	0.02	< .001	< .001	—					
Students per teacher	Pearson's r	0.19	-0.06	0.06	-0.01	-0.01	—				
	p-value	0.03	0.50	0.51	0.95	0.91	—				
Math Prof	Pearson's r	0.09	-0.15	0.17	0.11	0.03	0.39	—			
	p-value	0.38	0.15	0.11	0.27	0.75	< .001	—			
Reading Prof	Pearson's r	0.12	-0.14	0.18	0.11	0.05	0.31	0.93	—		
	p-value	0.26	0.18	0.08	0.27	0.66	0.00	< .001	—		
Econ Disadv	Pearson's r	0.15	-0.11	0.10	0.15	0.04	0.29	0.71	0.76	—	
	p-value	0.12	0.28	0.34	0.12	0.73	0.00	< .001	< .001	—	
Self-derivation	Pearson's r	0.17	0.13	0.35	0.15	0.34	-0.01	0.05	0.07	0.02	—
	p-value	0.03	0.10	< .001	0.05	< .001	0.95	0.59	0.47	0.85	—

Table 3: Factor loadings from exploratory factor analysis

	Factor 1 (School)	**Factor 2 (Cognitive)**	**Uniqueness**
Reading Prof	0.963		0.061
Math Prof	0.948		0.094
Econ Disadv	0.764		0.406
Vis-Aud Learning		0.736	0.462
Verbal Comp		0.698	0.5
Visualization		0.495	0.745
SES			0.895
Students per teacher			0.861
Extracurricular activities			0.919

Note. Applied rotation method is promax.

Table 4: R-Squared values from each SEM model

R-Squared

	R^2	
	Model 1	**Model 2**
Verbal Comp	0.611	0.557
Visualization	0.206	0.218
Visual-Auditory Learning	0.495	0.5
Math Prof	0.884	0.902
Reading Prof	0.989	0.973
Econ Disadv	0.645	0.844
Self-derivation	0.219	0.215
Students per teacher		0.145
SES		0.02
Extracurricular activities		0.002

Table 5: Factor loadings for model 1

Factor Loadings

Latent	Indicator	Estimate	Std. Error	z-value	p	95% Confidence Interval Lower	95% Confidence Interval Upper
Cognitive	Verbal Comp	1	0			1	1
	Visualization	0.704	0.155	4.542	< .001	0.4	1.008
	Vis-aud learn	1.095	0.192	5.697	< .001	0.718	1.472
School	Math Prof	1	0			1	1
	Reading Prof	0.942	0.045	21.027	< .001	0.854	1.03
	Econ Disadv	1.098	0.092	11.932	< .001	0.918	1.278

Table 6: Regression coefficients for SEM Model 1

Regression coefficients

						95% Confidence Interval	
Predictor	**Outcome**	**Estimate**	**Std. Error**	**z-value**	**p**	**Lower**	**Upper**
Cognitive	Self-derivation	0.011	0.002	4.543	< .001	0.006	0.016
School	Self-derivation	-5.41E-04	0.001	-0.472	0.637	-0.003	0.002
SES	Self-derivation	0.003	0.003	1.089	0.276	-0.003	0.01

Figure 1: Path diagram from exploratory factor analysis

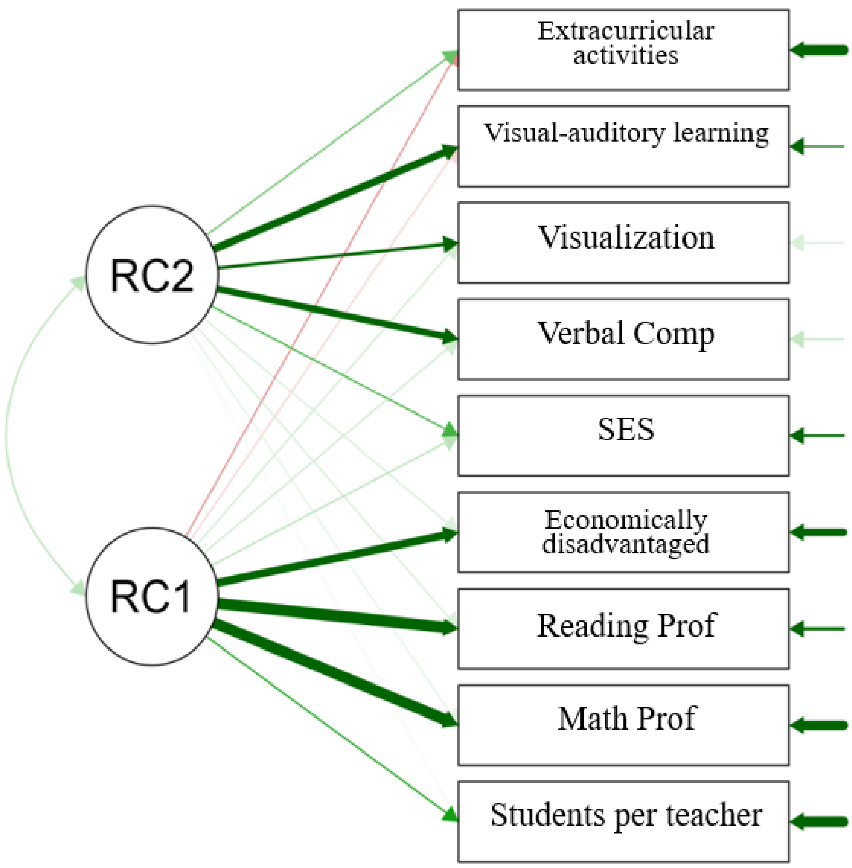
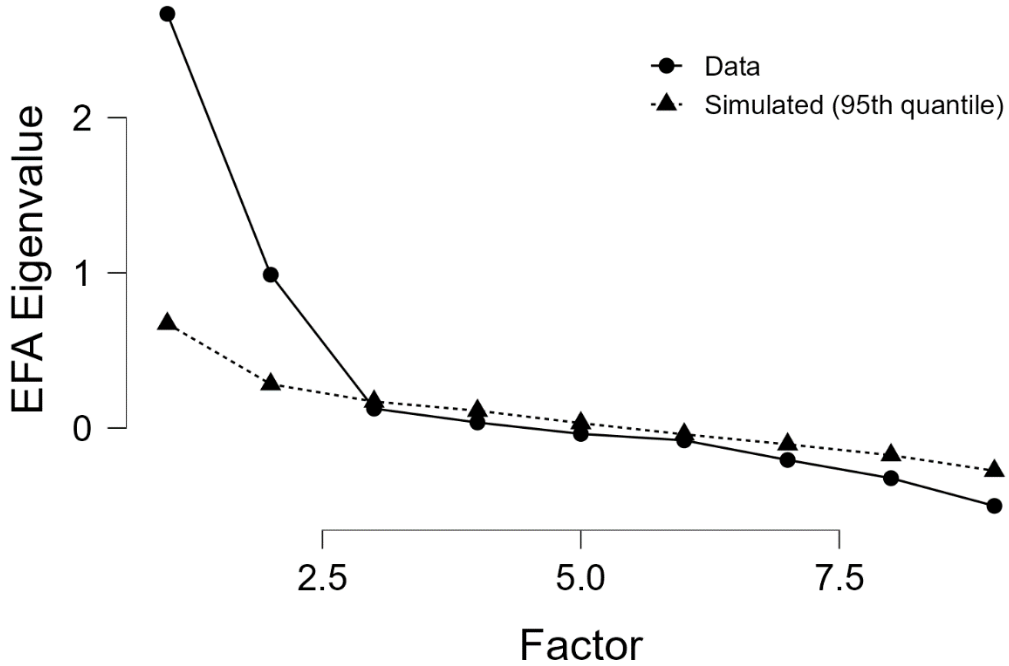
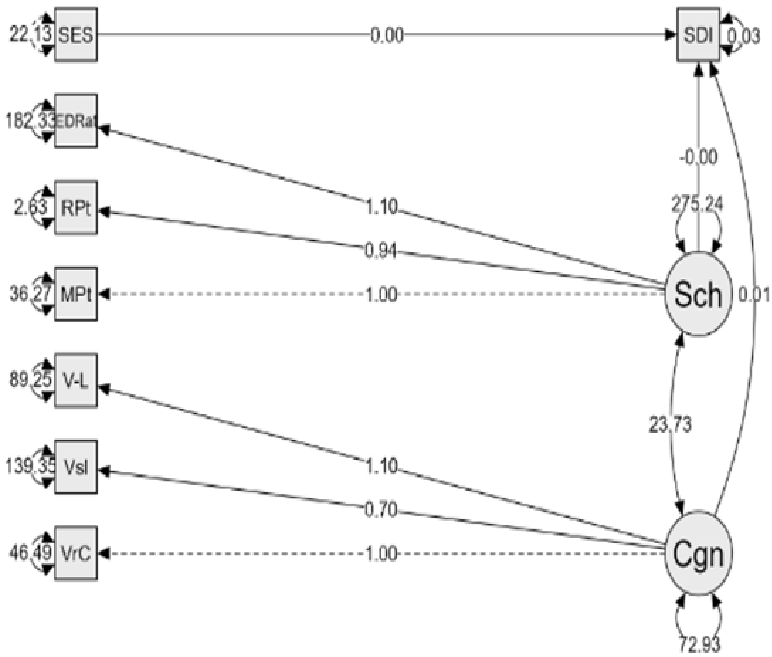


Figure 2: Scree plot of SEM model 1



Note. The number of points above the "leveling off" of the y-axis indicates how many factors should be retained. In this case, using the data line (circles) we determined there are two points before the y-axis (Eigenvalues) levels off. Therefore, we retained two factors or latent constructs.

Figure 3: Path diagram for SEM model 1



Key:
SES = Socioeconomic status
EDRa = Economically disadvantaged
RPt = Reading proficiency
MPt = Math proficiency
V-L = Visual-auditory learning
Vsl = Visualization
VrC = Verbal comprehension
SDI = Self-derivation performance
Sch = School latent construct
Cgn = Cognitive latent construct