Title
Cognition and Character: Measuring and Assessing Intellectual Development in Higher Education

Author
Orona, Gabe

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Cognition and Character: Measuring and Assessing Intellectual Development in Higher Education

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Gabe Avakian Orona

Dissertation Committee:
Professor Richard Arum, Chair
Distinguished Professor Jacque Eccles
Associate Professor Di Xu

2022

DEDICATION

I dedicate this to my family and friends.

To my beautiful fiancé: You are the kindest, most humble human being I have ever known. Your support and understanding during the challenging months leading up to the completion of this dissertation cannot be understated. I'm truly blessed to have you.

To my parents: You provided support in one form or another my entire life. I'm so grateful for you both.

To Uncle Jerry: You always encouraged my brothers and I to be thinkers; to not be lazy with our minds and whatever we do, do it thoroughly.

To my siblings: Anthony, Tommy, and Alana—I always want the best in each of your lives.

"Wisdom is the principal thing; Therefore get wisdom. And in all your getting, get understanding."

Proverbs 4:7

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge my primary advisor and dissertation chair, Richard Arum, for his constant support and guidance throughout this journey. I'm honored, impressed, and humbled that, during his service as Dean of the School of Education, Richard always made time to meet with and advise me. He has always been supportive of my research and has gone out of his way to ensure that I had the necessary training, resources, and funding to pursue my intellectual interests. His guidance was often necessarily pointed: reining in my all-to-frequent tangents and rabbit trails. He has been the quintessential opportunity provider, allowing for intellectual exploration and fulfillment. And for that, I'm very grateful.

Distinguished Professor Jacque Eccles has also had a profound impact on my thinking, and most of all the very direction of my research. It was Jacque who challenged me to move beyond superficial frameworks and deeply engage the cognitive development literature. I found this to be among the more difficult yet rewarding experiences of my PhD training, re-shaping my research identity. Her impact is sure to be lasting. I would also like to thank Di Xu, the final member on my committee, who has provided invaluable guidance on topics as broad as methodology, journal submissions, navigating complex data collection procedures, and more. She also provided funding for additional psychometric training and pushed me towards excellence in my work. Di's enthusiasm and productivity are inspiring; it continues to be a source of motivation.

I would also like to point to two informal guides who have made a special mark on my time at UCI: Greg Duncan and Mark Warschauer. Both treated me as if I was their student, providing time, advice, walks, and phone calls. My walk-talks with Greg are among the most memorable things about UCI. Mark's deep insights regarding job market strategies and affable personality made for some unrivaled highlights.

There are many peers to thank, as well. Sabrina Solanki, Remy Pages, and Sharin Jacob were constant friendly faces and very helpful in my barrage of questions about milestone requirements. My friends, colleagues, and world-class philosophers Daniel Herrmann, Aydin Mohseni, and Gerard Rothfus have had a tremendous impact on my thinking and approach to statistical modeling. The after-hour, late-night conversations about Bayesian epistemology, hypothesis testing, and probability theory are the stuff of graduate school legend. I am deeply grateful for each of them.

Additionally, I would like to thank Professor Jeff Rouder for showing me his tips and tricks (e.g., writing APA papers in Rmarkdown, etc.) for being a prolific scholar and statistical modeler. Additionally, I would like to thank Informa UK Limited (trading as Taylor & Francis Group) for permission to use copyrighted material for Chapter 1; and SAGE Publishing for permission to use copyrighted material for Chapter 2. The source of this permission is listed on the publishers' websites and confirmed via the editors.

VITAE

EDUCATION

2017-Present	**Ph.D. (expected 4/2022)**	Education
		Concentration in Measurement and Assessment in Higher Education
2015	**M.P.H.**	Master of Public Health, California State University, Fullerton
		Concentration in Quantitative Methods
2013	**B.S.**	Health Science, California State University Fullerton

PROFESSIONAL EXPERIENCE

2020	Educational Testing Service (ETS) Intern, National Assessment of Educational Progress (NAEP)
2018-Present	Graduate Student Researcher, Next Generation Undergraduate Success Measurement Project (Carnegie Mellon Foundation Project), School of Education, University of California Irvine
2015-2017	Research Analyst, Institutional Research Planning and Effectiveness, Citrus College
Spring 2015	Epidemiology Intern, Epidemiology and Assessment, Orange County Health Care Agency, Santa Ana, CA
2013-2015	Graduate Research Assistant, Department of Health Science, California State University, Fullerton
2013-2014	Clinical Trial Research Assistant, Neurology Center of North Orange County, Fullerton, CA
2008-2013	Imaging Assistant, Department of Radiology, Inter-Community Hospital, Covina, CA

Consultantships

2021	Student Experience in the Research University, UC Berkley; Cognitive Interviews, presentation, analysis
2016-2018	California Community College Chancellor's Office, Institutional Effectiveness Partnership Initiative; Program Evaluation and Measurement
2016-2017	Mt. Sac; Statistical Analysis

HONORS AND AWARDS

2021	Society for Research on Educational Effectiveness (SREE) Scholar, $8,000 Funding provided by the *Bill & Melinda Gates Foundation.* Project: *Aligning skills, competency frameworks, and learning outcomes*
2020-2021	David P. Gardner Research Seminar Fellowship on Higher Education, Center for Studies in Higher Education, University of California, Berkeley
2020	Educational Testing Service (ETS) Intern, National Assessment of Educational Progress (NAEP)
2017	Teaching and Learning Fellowship, University of California, Irvine
2017	UC Provost Fellowship, University of California, Irvine
2017	Diversity Fellowship, University of California, Irvine
2015	Outstanding Graduate Student Scholarship Award, California State University, Fullerton
2015	Departmental tutoring award
2014	Independent Research Scholarship (Clinical Trial), Department of Health Science
2013	Eta Sigma Gamma, National Health Education Honorary
2011-2012	Dean's List, California State University, Fullerton

RESEARCH INTERESTS

Measurement and Assessment in Higher Education

Statistical Modeling

Cognitive Development

Educational Psychology

PUBLICATIONS

1. **Orona, G. A.**, Li, Q., McPartlan, P., Bartek, C., & Xu, D. (2022). What predicts the use of interaction-oriented pedagogies? The role of self-efficacy, motivation, and employment stability. *Computers & Education*, 104498.

2. Arum, R., Eccles, J. S., Heckhausen, J., **Orona, G. A.,** von Keyserlingk, L., Wegemer, C. M., ... & Yamaguchi-Pedroza, K. (2021). A Framework for Measuring Undergraduate Learning and Growth. *Change: The Magazine of Higher Learning*, *53*(6), 51-59.

3. Fischer, C., Baker, R., Li, Q., **Orona, G. A.,** & Warschauer, M. (2021). Increasing success in higher education: The relationships of online course taking with college completion and time-to-degree. *Educational Evaluation and Policy Analysis*, 01623737211055768.

4. **Orona, G. A.** (2021). Gotta know why! Preliminary evidence supporting a theory of virtue learning as applied to intellectual curiosity. *Theory and Research in Education*, *19*(3), 279-295.

5. **Orona, G. A.**, & Pritchard, D. (2021). Inculcating curiosity: pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment & Evaluation in Higher Education*, 1-15.

6. **Orona, G. A.** (2021). Philosophy's rematch: A new conceptualization of the study of higher education. *Arts and Humanities in Higher Education*, 14740222211002967.

7. **Orona, G. A.** (2021). Broken promises? Examining the effectiveness of promising practices in STEM lectures by student subgroups. *Innovative Higher Education*, *46*(2), 223-239. https://doi.org/10.1007/s10755-020-09536-4

8. Baker, R., & **Orona, G. A.** (2020). Gender and racial differences in awareness and consideration of curricular programs: Exploring a multistage model of major choice. *AERA Open*, *6*(3), 2332858420937023.

9. Fagioli, L. P., Baker, R., & **Orona, G. A.** (2020). The role of non-cognitive variables in identifying community college students in need of targeted supports. *Research in Higher Education*, 1-39.

10. Piazza, J. R., Dmitrieva, N. O., Charles, S. T., Almeida, D. M., & **Orona, G. A.** (2018). Diurnal cortisol profiles, inflammation, and functional limitations in aging: Findings from the MIDUS study. *Health Psychology*, *37*(9), 839.

Manuscripts In Peer-Review

1. Fischer, C., Baker, R., Li, Q., **Orona, G. A.**, & Warschauer, M. (in review). Salient syllabi: Examining design characteristics of science online courses in higher education. *PLOS ONE*

2. Xu, D., **Orona, G.A.,** Li, Q. Interaction-oriented practices show promise in enhancing performance and narrowing racial achievement gaps in college online courses. *PLOS ONE*

Reports

1. Arum, R., Eccles, J. S., Heckhausen, J., **Orona, G. A.,** von Keyserlingk, L., Wegemer, C. M., ... & Yamaguchi-Pedroza, K. (2021). Ensuring a more equitable future: Assessing student learning and growth in higher education. Retrieved from: https://www.postsecondaryvalue.org/wp-content/uploads/2021/05/PVC-Arum-FINAL.pdf

2. **Orona, G. A.**, Li, Q., McPartlan, P., Bartek, C., Xu, D. (2020). Surveying online interaction: Relating college instructor characteristics and perceptions to online practices. University of California, Irvine. Online Learning Research Center. Retrieved from: https://www.olrc.us/student--instructor-perceptions.html

Published Data

Warschauer, M., Reimer, L., Denaro, K., **Orona, G.A.,** Schenke, K., Nguyen, T., Niili, A., Solanki, S. Tamara, T. (2021), Evaluating Promising Practices in Undergraduate STEM Lecture Courses, Dryad, Dataset, https://doi.org/10.7280/D11M5Q

RESEARCH PRESENTATIONS AND CONFERENCES

Orona, G.A., Eccles, J., Arum, R., & Pages, R. (2021, April). *Confirming Bias in Higher Education: Introducing and Validating a Confirmation Bias Performance Assessment*. NCME Annual Conference. Virtual Meeting.

Orona, G.A. & Pritchard, D. (2021, April). *Validating a Liberal Arts Inclination Scale: Applications in Classical and Modern Test Theory*. NCME Annual Conference. Virtual Meeting.

Orona, G.A. & Pritchard, D. (2021, April). *Inculcating intellectual character: Pilot results of an online module to enhance undergraduate intellectual virtue at university*. NCME Annual Conference. Virtual Meeting.

Orona, G.A. (2020, Apr 17 - 21) *Policy or Pedagogy? Do Promising Practices and Teaching-Focused Faculty Have an Effect on Student Outcomes?* [Paper Session]. AERA Annual Meeting San Francisco, CA http://tinyurl.com/tlzqjgl (Conference Canceled)

Orona, G.A. (2019, November). Policy or Pedagogy? *Evaluating the Interaction between Instructor Incentives and Teaching Practices on Student Outcomes*. Paper session presented at Association for Public Policy Analysis and Management Annual Conference, Denver, Colorado.

Orona, G.A., Vu, V., Warshauer, M. (2019, April). *Into the Wild: An Empirically Based Framework for Understanding STEM Pedagogy in University Education*. Paper session presented at the American Educational Research Association Annual Conference, Toronto, Canada.

Fagioli, L., Baker, R., & **Orona, G.A.** (2019, April). *The Role of Non-Cognitive Variables in Identifying Community College Students in Need of Targeted Supports*. Paper session presented at Association for Public Policy Analysis and Management Annual Conference, Student Conference, Irvine, California.

Orona, G.A., Baker, R., Denaro, K. (2018, November). U*sing Multiple Correspondence Analysis to Understand the Implementation of Promising Practices Across Instructor Characteristics in Large Undergraduate STEM Lecture Courses.* Panel session presented at Association for the Study of Higher Education. Tampa, Florida.

Fagioli, L., Baker, R., & **Orona, G.A.** (2018, November). *The Role of Non-Cognitive Variables in Identifying Community College Students in Need of Targeted Supports*. Paper session presented at Association for Public Policy Analysis and Management Annual Conference, Washington D.C.

Fagioli, L., Baker, R., & **Orona, G.A.** (2018, June). *The Role of Non-Cognitive Variables in Identifying Community College Students in Need of Targeted Supports*. Paper session presented at Community College 2 PhD Annual Conference, UC Los Angeles

Fischer, C., **Orona, G.A.,** McPartlan, P., & Yu, R. (2018, May). *Summer Sessions at a Large Public Research University: Evaluating Student Success in Undergraduate Online Courses*. Paper session presented at Digital Learning in the Humanities and Beyond: A UC Irvine-Tel Aviv University Symposium, Irvine

Orona, G.A. (2016, November). *Leveraging the Power of Regression Discontinuity Designs for Program Evaluation: An Institutional Simulation Using R*. Concurrent session presented at the annual conference of the California Association for Institutional Research, Los Angeles

Orona, G.A., Piazza, J.R., Zarit, S. Almeida, D.M. (2016, November). *The Role of Event Centrality in Predicting Diurnal Cortisol*. Symposium paper presented at the annual conference of the Gerontological Society of America, New Orleans

Huang, Y., **Orona, G.A.** & Stallard, C. (2016, April). *Dream Big: An Innovative, Unorthodox Strategy to Strategic Planning*. Presented at the annual Research and Planning Conference, San Diego

Orona, G.A. & Piazza J.R. (2015, May). *Acute Stressor Reactivity, Chronic Strain, and Cortisol: Examining the Physiopathologic Link*. Symposium paper presented to the Department of Health Science at the Masters of Public Health Research Symposium, California State University, Fullerton

Orona, G.A. & Piazza J.R. (2015, February). *Acute Stressor Reactivity, Chronic Strain, and Cortisol: Examining the Physiopathologic Link*. Symposium paper presented at the annual Student Research Competition, California State University, Fullerton

Orona, G.A., Piazza J.R., Stawski, R.S., & Almeida, D.M. (2014, November). *Event Centrality and Daily Well-Being: The Synergistic Effect of Chronic Strains and Daily Hassles*. Symposium paper presented at the annual conference of the Gerontological Society of America, Washington D.C.

Invited Talks/Guest Lectures

2021-March	"A Tale of Two Scales: The Pragmatic Benefit of The Construct-Modeling Approach Over The Usual Story In The Validation Of A Liberal Arts Inclination Instrument", Gardner Seminar, University of California, Berkeley
2020-July	"Who Does What? Utilizing process data to understand the relationship between problem-solving strategies and math ability", NAEP Internship, Educational Testing Service
2020-July	"What is College Worth?", Postsecondary Value Commission
2019- February	"Studies in Higher Education: Maximizing Learning through Policy and Pedagogy", Social Dynamics, UCI
2017- June	"Program Evaluation in Institutional Research", Regional Research and Planning CAMP meeting
2016- September	"Research Design", Undergraduate Research Methods, Citrus College
2016- June	"Regression Discontinuity Designs using R", Regional Research and Planning CAMP meeting
2015- April	"Quantitative Research Methods, Non-Experimental Study Designs", Undergraduate Research Methods, California State University Fullerton
2015- April	"Quantitative Research Methods, Non-Experimental Study Designs", Undergraduate Research Methods, California State University Fullerton
2015- February	"Epidemiology Study Design in Practice", Advanced Epidemiology, California State University, Fullerton

TEACHING EXPERIENCE

2021	Teachers Assistant: EDUC 15 LEC A: Statistics for Education Research, UC Irvine
2019	Statistics in R Graduate Series, "Factor Analysis", School of Education, UC Irvine

2014-2015	Statistics Tutor, Undergraduate Statistics, Department of Health Science
2014-2015	Epidemiology Tutor, Undergraduate Epidemiology, Department of Health Science

QUANTITATIVE SKILLS

Statistical Software:	R, Stata, SPSS, Mplus
Statistical Training and Coursework:	Todd Little's StatsCamp (2021): Bayesian Structural Equation Modeling; NCME: Bayesian Psychometric Modeling; Todd Little's StatsCamp (2019): Psychometrics; UCSB's MethodsU: Measurement; Statistics for Social Sciences; Advanced Statistics in the Health Sciences; Advanced Psychological Statistics; Educational, Social, and Behavioral Statistics; Longitudinal Structural Equation Modeling; Applied Regression; Data Analysis in Education; Causal Analysis; Psychometrics

RESEARCH GRANT AFFILIATIONS

Next Generation Undergraduate Success Measurement Project
Funding agency: The Andrew W. Mellon Foundation
Awarded amount: $1,115,000
Award number: #1806-05902
Role on project: Graduate Student Researcher
Grant awarded to Richard Arum (PI, University of California, Irvine) and Michael Dennin (Co-PI, University of California, Irvine)

EDITORIAL SERVICE

2014-2015	Reviewer: *California Journal of Health Promotion*

DEPARTMENTAL SERVICE AND VOLUNTEER

2021	Student Representative: Higher Education Faculty Hire Search Committee, University of California, Irvine
2018-2019	Educational Policy and Social Context (EPSC) Lab Coordinator, University of California, Irvine
2017-2019	Associated Doctoral Students in Education, Cohort Representative, University of California, Irvine
2015-2017	Student Equity Plan Evaluation Committee, Citrus College
2015-2017	STEM Evaluation Committee, Citrus College
2013-2015	Graduate Committee, Master of Public Health Representative, California State University, Fullerton
2013-2014	Research Assistant, Orange County Health Initiative, Orange County Health Care Agency

2012-2014 Research Assistant, Fibromyalgia and Chronic Pain Center, California State University, Fullerton

PROFESSIONAL SOCIETIES AND AFFILIATIONS

2020-Present	National Council on Measurement in Education
2018-Present	American Educational Research Association
2015-2018	The Research & Planning Group for California Community Colleges

REFERENCES

Richard Arum, Ph.D.
Distinguished Professor/Dean
School of Education
University of California, Irvine

Jacque Eccles, Ph.D.
Distinguished Professor
School of Education
University of California, Irvine

Andrew Maul, Ph.D.
Associate Professor
The Gevirtz School Education
University of California, Santa Barbara

Di Xu, Ph.D.
Assistant Professor
School of Education
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Cognition and Character: Measuring and Assessing Intellectual Development in Higher Education

by

Gabe Avakian Orona

Doctor of Philosophy in Education

University of California, Irvine, 2022

Professor Richard Arum, Chair

The value of higher education, and particularly the four-year undergraduate venture influenced from the liberal arts tradition, is largely framed around the impact it has on a broad range of skills and dispositions that serve to enhance human flourishing. Employers and society more broadly can benefit from individuals who are committed to careful thinking, and who are adept in navigating the complex and oft-confused torrent of information presented in everyday life. Among the most discussed attributes are critical thinking, perspective-taking and intellectual virtues, such as curiosity, open-mindedness, and humility. In response, this dissertation investigates undergraduate experiences that lead to the development of such desired qualities. The first chapter introduces and evaluates a pilot intervention to inculcate curiosity in university students, while examining the level of student satisfaction across important subgroups (N = 202). The second chapter, directly building off the first, tests a general theory of moral virtue development as applied to intellectual virtue by testing the mechanisms by which curiosity develops (N = 202). The third chapter examines the longstanding liberal arts notion of course-taking breadth on the formation of complex reasoning skills, situating the research within the adult cognitive development literature (N = 260). Results suggest that these undergraduate experiences are beneficial to the development of higher-order skills, though differences across cognitive outcomes are noticeable. Moreover, the magnitude of development shows promises for potential ways of scaffolding student skills. Lastly, while this research highlighted and was sensitive to issues relating to sound statistical practice (e.g., fitting many models and reporting all tests, employing Bayesian statistics, and directed acyclic graphs), more robust study design features should be implemented in the future.

INTRODUCTION

In the widely influential *Academically Adrift*, Arum and Roksa (2011) investigated undergraduate student learning by collecting empirical data on student involvements, study habits, and performance. In short, their analysis revealed two harrowing dimensions of the undergraduate student experience: (1) Students exert minimal time and energy into their studies, and (2) Students 'critical thinking and writing abilities improve marginally—if at all—during their time in college, bespeaking limited learning.

While critiques of Arum and Roksa's (2011) study include: the assessment of a general as opposed to a domain-specific skill(s), the possibility of measuring situational performance over actual competence, and other statistical issues, the impact of this work cannot be overstated. In recent years, discussion around the value of college-going for the development of skills and abilities has garnered increasing attention from researchers and policymakers alike. Voluntary reforms have been underway to improve measures of student learning, and a host of learning outcome initiatives have appeared.

Some of these developments include the Degree Qualifications Project and (Arum, Roksa, & Cook, 2016), general ability assessments, such as the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics developed by the Association of American Colleges and Universities (AAC&U, 2011), departmental and institutional level metrics produced by the National Institute of Learning Outcomes Assessment (NILOA, 2011), and, more recently, subject-matter concepts and competencies to improve the quality of student learning spearheaded by the Measuring College Learning (MCL) project (Arum, Roksa, & Cook, 2016).

Aside from methodological challenges facing these projects, other concerns include what kinds of competencies should be assessed, what should college graduates know, as well as how to resolve the long-standing tension between assessing general and discipline-specific knowledge and abilities in relation to student learning (Sparks, Song, Brantley, & Liu, 2014; Arum et al., 2016; Liu, Frankel, & Roohr, 2014)? Moreover, recent interest points towards measures that precede economic gains, ones that can be tied directly to the fundamental educational experience.

The belief that higher education does more—or should do more—than merely function as a screening check and training ground for employers is evident in that these measurement projects are seeking ways of expanding assessments beyond labor market outcomes. This belief is grounded in the long-standing tradition of liberal education. That general education produces a cultured, generally competent individual is part of philosophical orientation towards education that developed in the days of antiquity and is still promulgated today in nationwide higher education initiatives (AAC&U, 2011).

This review seeks to inform the discussion on measurement and assessment in higher education, and specifically undergraduate experiences intended to impact liberal arts outcomes. To that end, some history of the tradition is necessary. I provide a brief sketch of the development of the liberal arts conceptualizations over time up to modern day. Next, some of the major goals of liberal education are discussed in light of this history and with a view towards the latest developments in measurement. Additionally, for clarification purposes, I use liberal arts and liberal education synonymously as pertaining to theory emphasizing a well-rounded broad education aiming to have lasting effects on individuals' cognition and character. It should be noted that stricter definitions sometimes differentiate the former as only pertaining to the original, ancient curriculum (AAC&U, 2011).

The theory of a liberal arts education, in many respects, can be understood as directly arising from Aristotle's philosophy of education and his theory of ethics. (Likely, it goes back to Plato and even further.) In the *Nicomochean Ethics*, Aristotle describes two parts of the soul: one irrational and the other rational, with the latter separating humans from animals and maintaining the capacity for the development of moral and intellectual virtue. The two sets of virtue—moral and intellectual—were—theorized to develop via different routes. While the moral virtues such as courage, generosity, temperance, and justice, to name a few, were proposed to develop by habituation and training from youth, the intellectual virtues were proposed to be acquired through teaching. Aristotle stipulated five intellectual virtues: *techne* (technical expertise), *phronesis* (practical wisdom), *nous* (insight/grasping first principles), *episteme* (scientific knowledge), and *sophia* (theoretical wisdom, which is *nous* plus *episteme*). Being a treatise on human happiness, in the *Nicomochean Ethics*, the best life is asserted to be the life of contemplation, where theoretical wisdom is perpetually reflecting on itself. As Aristotle put it, "…what is best and most pleasant for a given creature is that which is proper to it. Therefore for man, too, the best and the most pleasant life is the life of the intellect, since the intellect is in the fullest sense the man. So this life will also be the happiest."

Another dimension of Aristotle's philosophy of education is the role education plays in a society more broadly. That is, different levels of education should exist for different needs within the state. The highest learning involving contemplation of first principles was proposed to be reserved for free citizens of wealthy backgrounds, while technical education, learning a craft or skill for practical purposes, was to best be left for those less advantaged at birth. Clearly, this is an excessively aristocratic, outdated, and self-serving viewpoint, but the idea that different forms of education are required for society's proper functioning and for the prosperity of the state is still evident today in contemporary federal calls for enhancing science, technology, engineering, and mathematics (STEM) education. Like today, Aristotelian purposes of education include both an individual epistemic dimension alongside a social and political one.

Centuries later, an important development in the liberal arts tradition took place in the Middle Ages, where Aristotle—along with Roman philosophers such as Cicero—were rediscovered and studied by Christian thinkers including St. Augustine and St. Aquinas, among others (McCaughey, 2019; Graham, 2002). Additionally, this is when the first European universities emerged, and when the structured curriculum consisting of the *trivium* (grammar, logic, rhetoric) and *quadrivium* (arithmetic, geometry, music, and astronomy) was formally introduced (Cronon, 1998). Here, disinterested inquiry is tied to the development of rationality alongside faith, understanding the Bible, and fulfilling one's calling unto God and for humanity. In *Augustine and Liberal Education*, Smith (2017, p. 209) summarizes the Christian perspective on liberal education:

Christ's sacrificial action restores our status as children of God. His redemptive defeat of death allows us the freedom to romp in the mind of God. In this sense, the Christian vision not only provides a psychological motive for liberal education by rendering the world investigable, it also provides a motive for learning by illuminating it's deeper meaning.

William Cronon (1998) describes how significant the Latin etymology is:

Liberal derives from the Latin liberalis, meaning "of or relating to the liberal arts," which in turn derives from the Latin word *liber*, meaning "free." But the word actually has much deeper roots, being akin to the Old English word *lëodan*, meaning "to grow," and *lëod*, meaning "people." It is also related to the Greek word *eleutheros*, meaning "free," and goes all the way back to the Sanskrit word *rodhati,* meaning "one climbs," "one grows." *Freedom* and *growth*: here, surely, are values that lie at the very core of what we mean when we speak of a liberal education. Liberal education is built on these values: it aspires to nurture the growth of human talent in the service of human freedom.

Ushering in the modern age, the Renaissance and the Scientific Revolution were periods of high creativity and intellectual breakthrough (Kato, 2016; Tachikawa, 2016). The Renaissance gave rise to humanistic education by placing salience on literature, poetry, art, rhetoric, and history, collectively known as the "humanities" (Kato, 2016; Tachikawa, 2016). Generally speaking, the humanities subsumed, somewhat, into the liberal arts curriculum and the two become closely associated (Kato, 2016; Tachikawa, 2016). Conversely, the Scientific Revolution sought to abandon these subjects in lieu of an emphasis on the new science. Tachikawa (2016) writes, "The new universal science which would transcend the largest religious conflict ever as the way to unity and truth entailed the decisive shift from Aristotelian logic to mathematical natural philosophy as the viable, central subject of liberal arts in the 17th century."

Although the Aristotelian preeminence of theoretical knowledge and the necessity of education for a vibrant nation-state were incorporated into the earliest conceptualizations of the first universities founded in the Middle Ages, and although the pressure to see the value of the humanities was already in full swing by the 17th century, it wasn't until the 19th century that the most popular thesis on liberal education would be written up and disseminated. This most profound and explicit formulation of a liberal arts education is arguably John Henry Cardinal Newman's (1852), *The Idea of a University*. What began as a series of lectures delivered at the commencing of Dublin's first Catholic university, the newfound rector, Newman, converted his notes into what is now understood to be a philosophical treatise spanning teaching and learning, curriculum, as well as the university's broader function.

Steeped in Aristotelian, Stoic, and Medieval philosophy, Newman defended the value of studying the classic subjects in an age of rapid advancement and industrial expansion. He candidly asserted that knowledge is its own end, worthy of pursuit irrespective of its instrumental value. However, his most memorable proposition confronts hyper-specialization and advocates for a general education that equips an individual for an array of tasks; he explicitly warned against producing individuals who perceived the world only through their discipline—individuals who attempt to fit every problem to their methodological repertoire and refuse to consider the lenses of others. For Newman, this was one of the great downfalls of a non-general education, and it is here where the notion of critical thinking and transferrable skills—alongside the attributes of Aristotelian moral and intellectual virtue—become so clearly associated with the theory of liberal education:

> His education is called 'liberal'. A habit of mind is formed which lasts through life, of which the attributes are, freedom, equitableness, calmness moderation, and wisdom; or what in a former Discourse I have ventured to call a philosophical habit. This then I would assign as the special fruit of the education furnished at a University, as contrasted with other places of teaching or modes of teaching. This is the main purpose of a University in its treatment of its students (Newman, 1852, p. 76-77).

Like Aristotle, Newman's philosophical habit of mind could not be the result of "mechanical" education, disconnected ideas, or hum-drum courses of instruction. Training and education were separate concepts, and their conflation was detrimental to cultivating the intellect to its fullest potential. Instead, developing this largeness of view is the concomitant effect of an arduous and continuous mental process of vetting new ideas against prior knowledge, an ardent disposition towards figuring out new combinations of things, and above all, apprehending the harmony between numerous facts, relations, and various types of knowledge. It's obvious, that for Newman, university education was an encompassing endeavor forever leaving a mark on those who embark on it. As quoted:

> We are instructed, for instance, in manual exercise, in the fine and useful arts, in trades, and in ways of business; for these are methods, which have little or no effect upon the mind itself, are contained in rules committed to memory, to tradition, or to use, and bear upon an end external to themselves. But education is a higher word; it implies an action upon our mental nature, and the formation of a character; it is something individual and permanent, and is commonly spoken of in connexion with religion and virtues (Newman, 1852, p. 86).

Newman's model for higher education, ironically, was more impactful in the structuring of U.S. universities than either England (his native country) or Dublin (where he was named university rector). Small liberal arts colleges and public institutions alike incorporated general education requirements, with an emphasis on literature and classical philosophical texts as part of well-rounded education for the whole person. But at the turn of the century, John Dewey's (1916) pragmatist influence on education—albeit less pertinent to postsecondary education—rejected the mind-body distinction, appealed heavily to empirical experimentation, and made frequent references to evolution as an analogy to developmental processes and learning. Dewey (1916) was a strong advocate of useful, practical knowledge and considered a myth the ancient Greek separation of "higher learning" from other forms of technical knowledge—something that seemed categorically at odds with Newman's notion of education as "a higher word".

Additionally, as technology continued to develop, "knowledge as its own end…" no longer held convincing force in light of nation facing the need to train WWII veterans and the broader population for jobs and positions requiring technical expertise (Pelikan, 1992; Graham, 2002; McCaughey, 2019). Addressing these developments, Robert M. Hutchins

(1952), president of the University of Chicago, initiated the Great Books Foundation—a program intended to preserve and promulgate the learning of classic Western philosophy and literature as a means to restore the idea of a liberal education. Hutchins (1952) likened the thread of textual discourse as engaging in an evolving human conversation with thinkers of the past—a way of gleaning the cumulative wisdom from the ages. His belief was that the "…liberal artist learns to read, write, speak, listen, understand, and think." Just as philosophy is considered to be unavoidable (even among those who attempt to do so), Hutchins (1952) suggests everyone is a liberal artist—some good, others bad. Accordingly, he saw the attributes associated with a "good one" as vital to democracy. And as to its relation to the movement towards STEM, he states, "Do science, technology, industrialization, and specialization render the Great Conversation irrelevant? We have seen that industrialization makes liberal education more necessary than ever and that the leisure it provides makes liberal education possible, for the first time, for everybody."

In recent years, equity concerns have grown to the forefront of postsecondary education. Ensuring an equitable society is now largely viewed as a key function of higher education and something worth continually striving towards. The idea that all members of society should have the right to access higher education is a largely held value. As such, postsecondary education is often believed to be the ideal lever by which disadvantaged and historically marginalized groups are able to secure social mobility (Pelikan, 1992). In *The Struggle for the American Curriculum, 1893-1958,* Herbert Kliebard (2004) chronicles competing philosophies of education, one of which—the social efficiency view—appeals to equity initiatives by suggesting that imparting knowledge and skills is the most efficient means to ensure a diverse, skilled workforce. For disadvantaged students, disinterested inquiry may still be a luxury they can't afford—they desire skills that will increase their chances of secure employment and high-paying careers that have, for various institutional, social, and cultural reasons, historically evaded them (Humphreys and Davenport, 2005).

In 2005, the Association of American Colleges and Universities (AAC&U, 2011) launched an intentional effort to describe and understand the value of a liberal education for the modern day (Humphreys & Davenport, 2005). In AAC&U's Liberal Education journal, an unequivocal shift in traditional thinking was identified: Liberal education is not a high, lofty pursuit for the wealthy in society, but an absolute necessity to prepare emerging professionals and citizens of all backgrounds for a perpetually complex and changing world. Thus, this 21st century conceptualization promulgates "practical liberal education" for all students. (It should be noted that the connection between workplace/societal needs is an idea traced back to antiquity and not newly introduced in the 21st century.) The campaign—Liberal Education and America's Promise: Excellence for Everyone as a Nation Goes to College (LEAP)—defined the liberal arts in the following way: "Liberal education is a philosophy of education that empowers individuals, liberates the mind from ignorance, and cultivates social responsibility. A liberal education comprises a curriculum that includes general education that provides students broad exposure to multiple disciplines and more in-depth study in at least one field or area of concentration" (Humphreys & Davenport, 2005).

However, LEAP leaders intended to not only promote the values of liberal arts education, but also to discover perceptions among students, faculty, and community and business leaders (Humphreys & Davenport, 2005). With this goal in mind, the AAC&U conducted eight focus groups across four different areas of the country with both high-school seniors and advanced college students at various institution types to gain student perceptions of the value of an undergraduate degree, generally, as well as liberal education, specifically. The results of these efforts catalogued three sets of student value rankings into a most, middle, and least valued set of outcomes (Humphreys & Davenport, 2005).

The students surveyed most valued work-related skills, such as time-management and teamwork. Other skills valued included business skills (e.g., leadership skills, exposure to the business world), alongside reasoning abilities such as critical thinking and analytic problem-solving. The least valued outcomes included, surprisingly, democratic and social outcomes such as sense of values, principles, ethics, tolerance and respect for other people, and appreciation of one's role as a citizen and an orientation toward public service (Humphreys & Davenport, 2005). The researchers noted that it wasn't that students did not value these qualities in and of themselves, rather, many of them did not see the relevance higher education has in inculcating these qualities; others perceived themselves as already adept in these areas.

As mentioned, the new liberal education is committed to equitable outcomes across sex, ethnicity, and socio-economic status (SES). What was once limited by class structure is now made available for all students, from all backgrounds. The focus on diversity has also concomitantly shifted the emphasis away from the curriculum and onto student experiences, both extra-curricular activities and pedagogical innovations. For instance, the WABASH National Study of Liberal Education emphasized a diversity of experiences, pedagogical practices, and other active participation concepts in the evaluation of student growth in college.

The practical turn hasn't taken off without some hinderances, though. As colleges and universities across the United States and abroad have accepted both the emphasis on practical STEM and non-STEM skills and general education requirements—a mandate handed down from the tradition of broad, well-rounded liberal education—the quality of both efforts appear compromised (Haberberger, 2018; Higgins, 2017).

The liberal arts notion of a broad education is most obviously identified in general education requirements, where university students must take courses outside their major. General education (GE) is sometimes criticized as a disconnected medley of courses, with no theoretical binding (Higgins, 2017). Indeed, a survey of 12,000 undergraduates revealed that 67% choose GE courses because it best fits their schedules (Seeley et al., 2018). Another recent survey administered to university students revealed that 72% preferred to take additional courses related to their major instead of GE requirements, and only 50% saw GE's as relevant to their major and future career (Thompson, et al., 2015). This is despite the fact that a whopping 70% understood that the *aim of GE* is to help student become more well-rounded and responsible individuals (Thompson, et al., 2015).

The brief historical description just provided is merely intended to provide an overview of the major Western epochs of thought associated with the liberal arts. The purpose was to gain sufficient working understanding of how these ideas manifest in the current rhetoric and operation in today's higher education context. While the account serves to provide a broad picture of the goals of liberal education, we turn our attention to two broad sets of intellectual outcomes. These include cognitive skills (e.g., critical thinking, etc.) and related dispositions (e.g., intellectual virtues).

University education—and specifically liberal arts education—has long been concerned with the development of general critical thinking and reasoning abilities. As Newman (1852) noted, "…a University, taken in its bare idea…educates the intellect to reason well in all matters, to reach out towards truth, and to grasp it." As part of the "Great Conversation," Hutchins (1952) describes the product of liberal education as an individual who can ascertain the fundamental premises, problems, and principles across disciplines and subject matter. This kind of general ability is bound by a general content and a general disposition towards problem-solving.

Critical thinking is, arguably, the quintessential higher education outcome (Hart Research Associates, 2016). In the early 1960's, the Watson-Glaser Critical Thinking Appraisal (WGCTA) was presented as a measure of students 'reasoning capabilities and continued to be used decades after its introduction (e.g., Ghanizadeh, 2017). Subsequent work ensued to define and operationalize critical thinking, such as Garrison's (1992) five-stage model that delineates a linear process consisting of elementary clarification, in-depth clarification, inference, judgement, and strategy formation. The test's salience has remained even when considering non-traditional/face-to-face settings, including those involving group and computer-assisted modalities (Newman, Webb, & Cochrane, 1995).

Often, critical thinking is also used as a larger umbrella term to include diverse forms of thinking aside from closed-form inferences and argument evaluation, such as metacognition and epistemic cognition (e.g., thinking about nature of knowing, the grounds for belief justification, and the subjectivity of knowledge claims). The 1970's ushered in one of the initial studies of this variety when Perry introduced his scheme of intellectual development, which is widely recognized as the pioneering work in the field of college student cognitive development. Perry (1970) conducted interviews with 140 undergraduates from Harvard and Radcliffe during the late 1950's and early 1960's, documenting their reflections in an open dialogue. His many interviews over the course of their college experience formed the basis for his nine-point developmental scheme which, at the time, was among the first to reveal how traditional college-age learners' thinking evolved from a dualistic/objective worldview to one where context and contingencies highlight the complexities with knowledge claims.

While studies have largely validated Perry's (1970) developmental positions (Erwin, 1983; Fago, 1995), the scheme has been indirectly confirmed with a plethora of model variants. For instance, in the 1980's, Basseches (1984) promoted *Dialectical Thinking*, a mode of reasoning where opposing views are reconciled through discourse. Also, Belenky et al. (1986) in *"Women's ways of Knowing"* describe how college females progress from received knowledge to an understanding of constructed knowledge. And Sinnott introduced (1981), formalized (1998), and expanded (2008) the theory of *Postformal Thought*, which outlines nine thinking positions, such as apprehending those situations and problems that lend themselves to multiple causes, solutions, and vantage points.

In the 1990's, Magolda Baxter (1992), in her *Epistemological Reflection Model*, delineates a pattern of growth beginning with absolute knowing (e.g., authorities provide certainty) to contextual knowing (e.g., contextual evidence is required for evaluation and finding solutions; no universal solutions). Chandler et al. (1990) researched *Epistemological Doubt*, which has three general steps closely mirroring Perry's scheme. Philosophically grounded in Dewey's (1933) notion of reflective thinking, and motivated by both the overemphasis on well-structured problems and the absence of epistemic assumptions in models of critical thinking, King and Kitchener (1994) describe the *Reflective Judgement Model* as, "…a developmental progression that occurs between childhood and adulthood in the ways that people understand the process of knowing and in the corresponding ways that they justify their beliefs about ill-structured problems."

The research on reflective judgement is unique in that an accompanying assessment procedure–the Reflective Judgement Interview (RJI)–was developed alongside the model and has been validated across several key criteria. The RJI has demonstrated good internal consistency across many studies (e.g., Brabeck, 1983; King, Kitchner and wood 1990), robust external correlations with other intellectual measures (Jensen, 1998), and clear developmental trajectories based on both cross-sectional and longitudinal data. For this latter point, as predicted by reflective judgement model, significant differences in mean RJI scores have been found across age categories and education levels.

While the early 2000's saw a few more model variants, such as Kuhn et al's. (2000) *Epistemological Understanding* model, by the 2010's most critical thinking research in higher education emphasized operationalizations resembling the Watson-Glaser assessments. Numerous exams (see Liu et al, 2014 for a comprehensive list)—orthogonal to any cognitive developmental models—have been developed, implemented, and evaluated in four-year university settings. Thus, the cognitive emphasis in higher education has largely returned to that of well-structured problem-solving.

Among the most widely cited is the collegiate learning assessment (CLA), which was developed as "a computer administered, open-ended (as opposed to multiple-choice) test of analytic reasoning, critical thinking, problem solving, and written communication skills" (Klein, Benjamin, Shavelson, & Bolus, 2007). Arum and Roksa (2011) utilized this measure in a nationally representative sample of 2,322 students who took the CLA in their first two years of college. Advancements in test design and technology have allowed for a revision to some of the mainstay assessments. Many studies are beginning to rely on the Educational Testing Service (ETS) designed Heighten Critical Thinking assessment to evaluate cognitive gains in college (e.g., Liu et al., 2016; Roohr et al., 2019).

Still, critical thinking is viewed as a multi-dimensional construct that may also have dispositional qualities. Hitchcock (2020), in the Stanford Encyclopedia of Philosophy, notes that critical thinking is not a pure "ability" measure, but one that has historically and philosophically involved habits of mind or intellectual virtues. As aforementioned, the concept of an intellectual virtue is directly traced to Aristotelian ethics and contributed to shaping the justification of a liberal arts education (Newman, 1852).

Generally, intellectual virtues are regarded as dispositions, traits and abilities constituting one's cognitive character (Zagzebski, 1996). With the advent of virtue epistemology, their relevance for education has seen a focused resurgence (Baehr, 2013; Pritchard, 2013). Arguments have been presented that a legitimate aim of education is to adapt pedagogy for the purpose of cultivating these desirable character traits in students; that is, educating for intellectual virtue (Baehr, 2013). Indeed, a recent development attempts to encourage educators to expand measures beyond standardized tests and performance metrics—and their association with economic prosperity—to include a holistic set of educational goods that relate to a life of flourishing (Brighouse, Ladd, Loeb, & Swift, 2018).

Intellectual virtue is a complex set of constructs that intertwine reasoning, character, and motives for actions. Although there remains philosophical discord as to whether cognitive faculties—like perception, memory, and reasoning—or character traits constitute intellectual virtue, a widely regarded work stipulates a distinction between cognitive virtues and skills (Zagzebski, 1996). Accordingly, the former constitutes such things as: open-mindedness, fairness in assessing evidence, intellectual humility, intellectual perseverance and thoroughness, ability to adapt one's intellect, among others that organize along the lines of malleable character traits as opposed to inherent or acquired intellectual prowess or ability (Zagzebski, 1996). An encompassing definition of an intellectually virtuous person as provided by Baehr (2013) includes "…one who desires and is committed to the pursuit of goods like knowledge, truth, and understanding."

In Character Strengths and Virtues: A Handbook and Classification, Peterson and Seligman (2004) set out to develop a comprehensive scale for the purpose of assessing virtues and character strengths over one's life span. With a positive psychology agenda, the impetus was to develop a reliable and valid diagnostic measure that is the positive counterpart to such diagnostic tools as the Diagnostic and Statistical Manual of Mental Disorders (DSM). The researchers incorporated texts from a variety of cultures, including Western and Eastern religious and philosophical traditions, to identity six cross-cutting virtues. The study resulted in the development of the Values in Action-Inventory of Strengths (VIA-IS) scale, which intends to measure the following virtues: courage, justice, humanity, temperance, transcendence, and wisdom.

Intending to validate the VIA-IS, Sing and Choubisa (2010) conducted an exploratory factor analysis on the VIA-IS and obtained a five-factor solution with different item loadings as stipulated by Peterson and Seligman (2004). The authors found the following factor structure: civic strengths, self-assurance strengths, interpersonal strengths, intellectual strengths, and theological strengths. Despite these efforts, virtue is notoriously difficult to measure (Kotzee, 2016), and new methods are being explored to inform educational practice (Jayawickreme, Meindl, Helzer, Furr, & Fleeson, 2014).

Aside from these general efforts to measure virtue, scales representing related constructs have been used and evaluated in higher education settings. For instance, the Need for Cognition (Cacioppo, Petty, & Feng Kao, 1984)–a variant of the virtue of intellectual curiosity (Lahroodi, 2007; Powell, Nettelbeck, & Burns, 2016)–has been widely used to evaluate the effects of liberal education (Pascarella, Wang, Trolian, & Blaich, 2013; Seifert, Goodman, King, & Baxter Magolda, 2010). Efforts have recently taken shape to introduce and validate a measure of intellectual humility, which is the characteristic of listening to others, avoiding overconfidence yet not undermining one's own knowledge (Krumrei-Mancuso & Rouse, 2016). Finally, a novel empirical data collection scheme to measure virtue reveals the building interest in collecting data for these constructs (Ng & Tay, 2020).

The goal of this three-study dissertation is to examine several college-going experiences that lead to liberal arts outcomes associated with cognition and character, each of which is sensitive to current statistical modeling perspectives given the advent of the replication crisis. Study 1 is a pilot evaluation of a novel online module to enhance intellectual virtue, focusing on the intellectual virtue of curiosity. The intervention is a set of online video recordings of professors and researchers discussing the importance and relevance of intellectual curiosity to their work and the work of prominent historical figures. The module also includes quizzes and work-along activities and encourages reflection of the material for personal implementation. The modules were informed by both virtue epistemic perspectives and interest theory (Hidi & Renniger. 2006). The module was uploaded to student course pages and available for extra credit.

The research design of study 1 was a one-group pretest-posttest design. Given this particular design's limitations for generating unbiased estimates of causal effects, I also employ prior information to regularize the estimates of interest using informed priors taken from the literature from similar studies. Following the procedures outlined by Gelman et al (2020), I utilize a Bayesian modeling strategy to examine the simple pre-post change.

Overall, I found that the intervention has a positive—though preliminary—association with increases in the Need for Cognition scale, Epistemic Curiosity scale, and two measures of the value for and knowledge of intellectual virtue. Furthermore, I found that several key student subgroups did not have significantly different affections when experiencing the intervention per equal levels of satisfaction across groups. I delineate the strategies for scaling up this pilot study and discuss implications for higher education.

Study 2 is a direct follow-up from study 1 and examines the links by which curiosity develops, as predicted by Besser's (2020) theory of virtue learning. Besser indicates three principal constructs that are necessary for the development of virtue. These include: (a) Knowing *what* virtue is; (b) Knowing *why* virtue is important and (c) Knowing *how* to implement virtue. In study 2, I use data collected originally for study 1 and investigate the *wha*t and *why* in Besser's model as it pertains to gains in intellectual curiosity. There is, to date and to my knowledge, no prior empirical study that explicitly employs a general theory of virtue learning to examine its applicability for the concept of intellectual virtue.

Study 2 is nonexperimental. Therefore, I employ several analytic approaches and do not *solely rely* on the use of *p*-values. I operationalize the variables in a variety of ways and report all tests. These tests include binary/descriptive associations, model comparison using Bayes Factors and other information criteria, and latent variable models.

Results of study 2 show—across all analytic approaches—that learning why intellectual virtue is important to one's education is a significant and moderate correlate of intellectual curiosity, while learning what is much smaller in magnitude. I discuss both the theoretical and practical implications of these results in the discussion section, highlighting how these results elucidate the relationship of moral and intellectual virtue, and how they may inform virtue education programs and interventions.

Study 3 examines the how out-of-major course breadth predicts complex cognitive reasoning, a construct formed by ill and well-structured problems. I also examine how individual attributes moderate the influence of course-breadth. Study 3 uses Fischer's skill theory, an enduring theory of cognitive development that's been used to synthesize models of intellectual development in college, to motivate and frame the analysis. In this way, study 3 engages with the cognitive psychology literature.

In study 3, a unique set of cognitive performance assessments spanning classical critical thinking tasks (e.g., argument evaluation), news source bias identification, and perspective-taking tasks, were deployed in fall 2019 and again in spring 2021. Combining data from administrative records and surveys, I construct a count of courses taken outside of students major and specify a series of Bayesian models with informative priors.

The results of study 3 show that the posterior distribution of the parameter for out-of-major course breadth is centered on a moderate-to-strong value (.17) and that that most of the probability mass is over positive parameter values. Additionally, intellectual curiosity appeared to have a noticeable moderating effect on outer breadth such that students high on curiosity and with many courses outside their major perform better than those that also have high outer breadth but are low on curiosity. However, this relationship is reversed for those with low outer breadth: curious students do worse than those less curious. For the global tendency to exert effort construct, the results did not provide evidence of an interaction effect. I discuss how the results relate to future research, educational programming, and the adult cognitive psychology literature in the discussion section.

All three studies intersect on the topic of cognition and character. They relate to issues concerning the hard and soft skills acquired during college-going that, in part, are used to evaluate the value of higher education. It's my hope that these studies can be fruitful lines of research that provide theoretical and practical advances in measuring and assessing development in postsecondary spaces, particularly those experiences associated with the idea of liberal education.

CHAPTER 1[1]

Inculcating Curiosity:

Pilot Results of an Online Module to Enhance Undergraduate Intellectual Virtue

1. Introduction

Recent educational work on student intellectual development has appealed to the framework provided by virtue epistemology, which is a theory emphasizing individual attributes in the belief formation process (Hyslop-Margison, 2003; Orona, 2021a). The intellectual virtues are admirable character traits of an individual that are geared towards specifically epistemic goods, like truth, knowledge, and understanding. Examples of intellectual virtues include open-mindedness, curiosity, intellectual courage, and intellectual humility. Like virtues more generally, the intellectual virtues involve both a motivational and a skill component (Zagzebski,1996). It has been argued that a fundamental epistemic goal of education is to develop the intellectual virtues (Pritchard 2013).

Consider the intellectual virtue of curiosity, for example, which involves being interested in acquiring new information and hence being willing to seek it out via questioning, observation, and so on. It is important to this intellectual virtue both that one is motivated by epistemic goods (as opposed, for instance, to someone who asks lots of question because they enjoy causing annoyance) and that one is skillful in how one

[1] This chapter is derived from an article published in *Assessment & Evaluation in Higher Education* on 24 Jun 2021, available online: https://doi.org/10.1080/02602938.2021.1919988. Full cite: Orona, G. A., & Pritchard, D. (2021). Inculcating curiosity: pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment & Evaluation in Higher Education*, 1-15.

undertakes one's inquiries (e.g., simply asking lots of irrelevant questions is not the manifestation of this intellectual virtue).[2]

The importance of the intellectual virtues is partly held to be due to how they enable subjects to be better positioned to pursue and acquire true beliefs (Dyer & Hall, 2019; Pritchard, 2019; Zagzebski, 1996). As such, scholars have begun to argue for the place of intellectual character development in the wider educational curriculum (Baehr, 2013, 2016; Pritchard, 2013; Barzilai & Chinn, 2018), and specifically in higher education (Battaly, 2006; Byerly, 2019; Dyer & Hall, 2019; Heersmink, 2018; Orona, 2021a; Schwartz; 2020). This line of inquiry centers on the question of how higher education practices, policies, and pedagogies could be devised to develop epistemically mindful individuals (Carter, 2017; Dyer & Hall, 2019).

However, most of the relevant higher education intellectual virtue scholarship remains non-empirical (e.g., Orona, 2021a; Schwartz; 2020). That is, there are few scientific studies of intellectual virtue among adults (e.g., Leary et al., 2017; McGrath et al., 2020; Zmigrod et al., 2019), and virtually none (to the authors' knowledge) that focus on inculcating these virtues among college-going students. Thus, with the current significance placed on the development of student dispositions, there is a critical need for research aimed at measuring and assessing intellectual virtue in higher education.

---

[2] For some contemporary treatments of the intellectual virtues, see Zagzebski (1996), Roberts & Wood (2007), Pritchard, Millar & Haddock (2010), and Baehr (2011). For an overview of the contemporary literature on this topic, see Battaly (2014) and Turri, Alfano & Greco (1999). For a recent discussion of the intellectual virtue of curiosity, see Ross (2020).

In this study, we introduce and evaluate a novel educational intervention designed with the explicit aim of enhancing students' intellectual virtues (in this pilot study, we focus on the virtue of intellectual curiosity). This offers a unique opportunity to evaluate the fidelity and preliminary effectiveness of developing the intellectual virtues within the specific context of university-level education, and thereby fills an important gap in higher education research. It also offers the further advantage of considering how educational interventions of this kind and at this level might function within an online setting. In evaluating the intellectual virtue intervention in the initial phase, we ask the following research questions (RQ):

- RQ1: To what extent are students satisfied with the intellectual virtue intervention module?
- RQ2: Do female, underrepresented minorities (URM), first-generation, and low-income students experience lower satisfaction with the intellectual virtue intervention than their counterparts?
- RQ3: To what extent does participation in the intellectual virtue intervention increase student learning gains in intellectual curiosity, knowledge of the virtues, and their perceived relevance to education?

1.1 Description of the module

The intervention is a short, online educational module embedded within courses as part of a larger university-wide project entitled 'Intellectual Virtues in the Curriculum' (IVC). The module has two broad components: (a) introducing and showing the significance of the concept of intellectual virtue, generally, and (b) an explicit emphasis on the virtue of intellectual curiosity. It contains a plethora of pedagogical design features, such as (pre-recorded) videos, engaging exercises, and information on of how the intellectual virtues relate to larger concepts in specific fields. Students received lectures, quizzes (which they were permitted to re-take), and activities on the IVC module to stimulate and potentially stabilize interest in intellectual virtues (i.e., attempting to trigger and maintain students' situational interest in intellectual virtues; Hidi & Renninger, 2006) and inculcate curiosity. Students were to complete the module at their own pace over the course of the term; however, in total, the allotted recorded lecture time—notwithstanding quizzes and activities—was about 2.5 hours of material, separated into 8 mini-modules. Due to space constraints, in Appendix A in the supplementary material, we present detailed information on the specific components of the module, the theory guiding the design of the pedagogical features, quiz items, and figures depicting the general format of the module.

2 Methods

2.1 Context

This study takes place at a large public research university located in southern California with a highly diverse student body. The IVC project is supported by internal funding that is aimed at both introducing and evaluating pedagogical innovation. Internal Review Board (IRB) approval was obtained during spring 2019, which approved the ethics of the student surveys and study intent. The IVC module was introduced across three different undergraduate courses: 'Introduction to Philosophy' (Philosophy1), 'Introduction to Ethics' (Philosophy4), and 'Frameworks for Professional Nursing Practice' (Nursing110W). All courses except the Philosophy1 were face-to-face courses. The IVC module is completely online and was made available to students as an extra-credit option for the three large lower-division courses in the fall 2019 quarter. There was no penalty for the students who chose to not participate. The module was accessible to students via a hyperlink tab on the course dashboard of the learning management system.

2.2 Procedure

Students who opted into the online module were administered a pretest and posttest survey. The surveys took approximately 10-15 minutes to complete. Informed consent was obtained prior to students completing any survey questions. All students were made aware prior to beginning the module of both the purpose of the study and that their voluntary participation can be redacted at any time. Data obtained from the surveys were later compiled with institutional records provided by the Teaching and Learning Research Center at the university.

2.3 Participants

There were 602 students with complete institutional data enrolled across the two philosophy and nursing courses. All students were provided with, and completed, a study information sheet that relayed the requirements of the study, the confidentiality agreement, as well as how the results would be used. 264 (44%) students opted into the IVC module. The completion rate was 77% (202 students completed the module). Table 1.1 displays the means and standard deviations (binary variables are interpreted as proportions) for students without missing data to compare variables across the full, participant, and completer samples. As can be seen, generally, the means and proportions of student characteristics, academic variables, and major clusters are comparable between the full course sample, participant sample, and the completer sample. All further analysis was performed on the completer sample (those with both pretest and posttest scores).

2.4 Measures

As mentioned, participants were asked to complete a pretest and posttest survey. Satisfaction measures were administered posttest only. Measures that were on both surveys included two curiosity constructs and two other subjective learning gain items, self-reported knowledge of the virtues, and their perceived relevance and importance to the students' education. To measure intellectual curiosity, we administered the 18-item need for cognition (Cacioppo, Petty, & Feng Kao, 1984) and 5-item epistemic curiosity (EC) sub-scale (Litman, 2008). We provide more details and background on these measures below.

Satisfaction measures. Student satisfaction is a common construct utilized in higher education research to evaluate and understand student learning experiences across a range of educational programs and pedagogical innovations (Armbruster et al., 2009; Doménech-Betoret et al., 2017; Lin & Chen, 2016; Overbaugh & Nickel, 2011). It is also widely used to evaluate institutional quality (Alves & Raposo, 2007; Santini et al., 2017). More recently, it is used as an outcome variable to understand the differential experiences of underrepresented minority students (Miller & Orsillo, 2020; Williams et al., 2018). In the present study, we combine these aspects and test whether satisfaction with the IVC module is biased against underrepresented student subgroups.

In this study and during the posttest, students were asked four satisfaction related questions: (S1): *Overall, how satisfied or dissatisfied are you with your experience with the intellectual virtue modules?* (S2): *How effective was the module at introducing you to the concept of intellectual virtues?* (S3): *In comparison to other extra credit opportunities you may have encountered in the past, did you find this module more beneficial in terms of gaining a quality learning experience?* (S4): *How likely are you to recommend this module to friends or colleagues?* Items S1 and S2 were positioned on a 5-point Likert scale, with higher numbers representing greater satisfaction. How likely students are to recommend the module to other students was measured on a slider ranging from 0 to 100, while how beneficial the module was in comparison to other extra-credit opportunities was measured on a 7-point Likert scale. These four questions (S1:S4) were subsequently used in RQ2 as indicators of a latent satisfaction factor.

Subjective knowledge and importance of intellectual virtue measures. The item relating to self-reported knowledge of the intellectual virtues was measured on a 3-point scale with 1 = 'I had never heard of them', 2 = 'I had heard of them, but I couldn't tell you what they are', and 3 = 'I had heard of them and could tell you what they are'. For perceived importance of the intellectual virtues to the students' education, response options included 1 = 'Don't know' 2 = 'Not important', 3 = 'Slightly important', 4 = 'Moderately important', 5 = 'Very important', 6 = 'Extremely important'.

Intellectual curiosity measures. The need for cognition scale (NFC) is an established measure defined as 'one's tendency to engage in and enjoy thinking' (Cacioppo & Petty, 1982, p. 130). The NFC has been used in numerous studies, including those showcasing its mediating role between personality and intelligence (Furnham & Thorne, 2013), as well as positive correlations with college grade-point average (GPA; Elias & Loomis, 2002), standardized test scores (Neigel, Behairy, & Szalma, 2017) , involvement in co-curricular activities during college (Wang, 2013), and skill acquisition (Day, Espejo, Kowollik, Boatman, & McEntire, 2007). Furthermore, international versions of the NFC have demonstrated similar positive correlations with college GPA (Salama-Younes, 2016). But more pertinent to the present study, NFC has been deemed a widely accepted measure of a desirable global trait that is expected to change over the course of undergraduate education at university. National studies investigating the effects of liberal arts education have used NFC as an outcome of college-going experiences, demonstrating its growth over time and with exposure to key instructional practices (Pascarella, Wang, Trolian, & Blaich, 2013).

The NFC was measured on a 5-point Likert scale ranging from 1 = 'Extremely uncharacteristic' to 5 = 'Extremely characteristic'. The pretest Cronbach's alpha for the NFC was α= .87. The posttest Cronbach's alpha for the NFC was α= .85.

Epistemic curiosity (EC) is a related construct defined as a 'desire for knowledge that motivates individuals to learn new ideas, eliminate information-gaps, and solve intellectual problems' (Litman, 2008, p. 1,586). In this study, we employ the five-item diverse (general) EC subscale. While not as heavily studied as NFC, EC has been related to grades (Eren & Coskun, 2016), among other interest or investment-trait variables (Litman & Spielberger, 2003).

The EC was measured on a 4-point Likert scale ranging from 1 = 'Almost never' to 4 = 'Almost always'. The pretest Cronbach's alpha for the EC was α= .88; the posttest Cronbach's alpha for the EC was α= .91. For both time points, the NFC and EC were moderately correlated. The pretest correlation was r = .50 and the posttest correlation was r = .54.

2.5 Data analysis

Analytic plan for RQ1. We answer RQ1 by descriptively examining the means and standard deviations for each of the four satisfaction measures. We look to see—given each response scale—whether satisfaction scores are above the mean response option.

Analytic plan for RQ2. We answer RQ2 with descriptive statistics and with a latent variable structural equation model (SEM). We regress a latent satisfaction construct—using the four satisfaction measures as indicators—on URM status, gender classification, first-generation and low-income status. Additionally, we control for Scholastic Aptitude Test (SAT) scores, previous school (either high school or transfer school) GPA, and whether the student is a science, technology, engineering, or math (STEM) major (coded as 1) or not (coded as 0). Due to low cell-size for majors, we only make STEM vs. non-STEM comparisons. Using conventional indices, we also evaluate model fit via evaluating the chi-square test (χ^2), confirmatory fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA).

Analytic plan for RQ3. For RQ3—our primary research question—we conduct a series of regressions on the four pre-post measures: the two curiosity measures, understanding of the virtues, and the perceived importance to education. For this question, we take a Bayesian approach to data analysis. The distinction with frequentist or classical statistics is primarily situated on the nature of probability, whereby Bayesian epistemologists argue for a personalistic view. On this account, probability is a belief of an individual, not an attribute of an event or object that is discovered in its long-run frequency. Indeed, there are many formal justifications and arguments supporting Bayesian probability in the philosophy of science literature (e.g., Savage, 1972), though they are too removed from the present purposes to recount here. However, as the philosophical distinction necessarily leads to different procedures and output in the application of common inferential tools, it seems necessary to provide a cursory explanation of Bayes' rule and a very brief summary of the benefits of Bayesian methods.

As a result of the large-scale endorsement among formal epistemologists, as well as increased computational power of modern software, Bayesian methods are slowly beginning to gain popularity in the social, behavioral, and educational sciences (Levy, 2016). Bayes' rule, from which the analytic approach gets its name, is described by the following:

$$Pr(\theta|y) = (Pr(y|\theta)Pr(\theta))/(Pr(y))$$

where θ is a hypothesis or parameter of interest, y is the data, $Pr(\theta)$ is the prior belief or prior probability of θ, $Pr(y|\theta)$ is the likelihood, $Pr(y)$ is the probability of the data, and $Pr(\theta|y)$ is the posterior distribution, representing the updated belief about θ, conditional on the data (y). The process of applying Bayes' rule is the precise and rationale reallocation of credence; the consequence is an updated belief. For example, an agent with a set of beliefs (prior) encounters evidence (data), those beliefs are updated according to Bayes' rule, reallocating probability mass from some proposition or range of parameter values to other propositions or parameter values, and hence a new probability distribution describing the agent's belief(s) emerges (posterior).

Unlike classical statistics, the primary output is not a point estimate and associated *p*-value, nor is it any other singular test statistic or widely accepted threshold. The primary output of a Bayesian analysis is the entire posterior distribution, which contains all information relevant to one's beliefs about the data, conditional on the data (e.g., Gelman & Shalizi, 2013; Morey, Romeijn, & Rouder, 2013; Kruschke & Liddell, 2018; McElreath, 2020). As priors are an integral aspect of Bayesian data analysis, we now turn to the construction of the priors as part of the analytic plan.

In constructing our prior distribution for the curiosity measures, we look to several higher education studies assessing change in NFC. Because there are currently, to the authors' knowledge, no comparative studies on the effects of similar online virtue modules to directly inform our prior distribution, our prior is constructed on the basis of the broader college-going literature. For instance, a large-scale study found that one year of college increases NFC by .11 standard deviation units (Seifert et al., 2010), while a .13 and .07 (Pascarella et al., 2013; Castle, 2014) standardized effect has been attributed to different liberal arts experiences. While the timeline for most of these studies focuses on a year or more worth of college, our study's timeline is one academic quarter. Naturally, this would decrease the magnitude of change anticipated in the curiosity measures. However, each of the independent variables in the listed studies represent broad college-going experiences that are not explicitly intended to inculcate curiosity. The IVC module, in contrast, is explicitly designed to develop students in intellectual virtue, and specifically curiosity, and therefore a larger effect than previous studies—despite the shorter time period—is anticipated, yet the limits of this effect are constrained by previous knowledge. Thus, we represent our prior beliefs about the (standardized) effect of the IVC module as normally distributed with a mean of .13 and a standard deviation of .1, reflecting our uncertainty about the module's association with the two curiosity constructs (NFC and EC).

Below we display the full model with a dummy-coded IVC variable (0 = pre-module and 1 = indicating post-module):

$$Y_t \sim Normal(\mu_t, \sigma)$$

$$\mu_t = \alpha + \beta_1(IVC)$$

$$\alpha \sim Normal(0, 2.5)$$

$$\beta_1 \sim Normal(.13, .1)$$

$$\sigma \sim Exponential(1)$$

where Y_t represents the outcome (NFC or EC) for every time-point t; α is the constant; β_1 is the parameter representing the effect of the dichotomous IVC variable comparing pre and post module scores (1 = post-module).

For the subjective gain measures, Y_t represents understanding of the virtues and the perceived importance to education, and the curiosity model structure remains largely intact, only with different priors. Here, we stipulate larger effects (and more uncertainty), since students are not routinely exposed to intellectual virtues or taught their connection to education (Hyslop-Margison, 2003; Baehr, 2016). Thus, we anticipate the effect of the IVC module (β_1) on these two outcomes to be normally distributed with a mean of 1 and standard deviation of .5.

3. Results

3.1 RQ1: Satisfaction and enjoyment of IV curriculum

Table 1.2 displays the overall statistics for each of the four satisfaction variables. We find that, for all four variables, satisfaction was well above the middle response point. Additionally, every subgroup examined displayed mean scores above the mid-point response of the scales. Finally, although only descriptive, the small standard deviations for each measure and across subgroups is also suggestive of the high-level of satisfaction with the IVC module.

3.2 RQ2: Equitable satisfaction and enjoyment of IV curriculum

Table 1.2 provides descriptive evidence showing that not only is satisfaction generally high, but that differential satisfaction across subgroups is also minimal. To formally test these differences, we present figure 1.1 showing the results of the SEM model, which fit the data well across multiple indices (χ^2 = 49.61 (26), p < .01, CFI = .941, TLI = .914, RMSEA = .066). Figure 1.1 displays two relevant features of the SEM model, starting from left to right: (a) The estimates for the variables used to predict the satisfaction with IVC module factor and (b) the factor loadings of the four satisfaction indicators (S1:S4) on one latent satisfaction factor. While the four satisfaction indicators each exhibited high loadings (well above .5), the latent satisfaction variable was not significantly predicted by any of the four subgroups of interest, after controlling for prior ability and major, $p > .05$.

3.3 RQ3: Learning gains

Table 1.3 displays the summary of the posterior distribution and convergence diagnostic of the IVC parameter for each of the four outcomes. Both subjective gain measures exhibited larger posterior means than the two curiosity measures. The importance of IV to education had the largest posterior mean with .75. Of the two curiosity measures, NFC had a larger posterior mean of .18. The convergence diagnostic shows a value of 1, indicating that posterior chains are well mixed for all models.

To visualize this updated belief about the IVC module's association with NFC, we turn to the top pane (a) of figure 1.2. Here, the left most distribution represents our prior belief about the effect of the IVC module on NFC before seeing the data which, as stipulated in the methods section, is represented as normally distributed with mean of .13 (indicated by the vertical line) and standard deviation of .1. The right most dotted outline represents the likelihood function for this parameter, also described as the information obtained from the data. The solid middle distribution is the posterior distribution, which here is clearly displayed as a product of the prior and likelihood. We see that the information obtained from the study data has shifted our belief about the IVC effect size to the right (i.e., placing greater density over larger positive values), though our prior belief is doing work in tempering the reallocation of credence over plausible parameter values.

Similarly, the bottom pane (b) of figure 1.2 shows the prior, posterior, and likelihood of IVC parameter (IVC effect size) in the EC model. Using the same prior as we did for the NFC model (M = .13, SD = .1), we see that the posterior shows only a marginal shift away from the prior, with all distribution outlines in the figure very close together. The posterior displays a tighter distribution, increasing our confidence that that the mean effect size is near .13.

The bottom pane (b) of figure 1.3 shows the prior and posterior distribution for the IVC parameter in this model. We see that our belief IVC effect size has also shifted downward, as we reallocate credence over plausible parameter values. Here, our posterior shifts away (downward) from our prior towards lower parameter values.

4. Discussion

In line with the growing interest in a virtue epistemic pedagogical framework and the newfound salience placed on non-academic metrics in higher education (e.g., Fagioli et al., 2020), the purpose of this pilot study was to provide preliminary evidence of the effectiveness of an online intellectual virtue module. As these pilot data indicate, students generally reported high levels of satisfaction with the IVC module. Additionally, our second set of findings suggest that a negative student experience with the module is likely not a function of membership in historically underrepresented and underprivileged groups. Given concerns with providing equitable educational programs and pedagogies that appeal to and resonate with a diverse student body (Orona, 2021b), we view these results as promising.

Our third set of findings suggest that students exhibit learning gains across all four measures. First, for the curiosity measures, the posterior means representing the standardized effects were larger than the point estimates reported in previous studies examining college or college-going experiences. Unlike other effects on NFC found in previous studies (e.g., Pascarella et al., 2013), we observed a shorter time period yet observed a larger effect size. But this was largely anticipated; previous studies were not tailored interventions to increase curiosity, but rather evaluations of the effects of broad college-going experiences.

Furthermore, students' self-reported claims about their knowledge and understanding of the value of the intellectual virtues sharply increased, showing preliminary evidence that students perceived the module as effective in initiating them to the intellectual virtues and drawing their connection to education. Though deploying objective longitudinal measures, including unique forms of performance assessments and in-depth interviews, constitutes a more robust scheme for assessing growth, self-report is commonly used to assess affective, behavioral, and cognitive learning gains in higher education research (Rogaten et al., 2019).

One obvious contention to this finding is that the current design does not control for time effects. That is, for example, we cannot rule out the possibility that the growth in curiosity is the result of natural maturation undergraduate students undergo during the quarter. Though, there are several features of the analysis that are important to highlight for the purpose of understanding this potential source of bias. Using the NFC findings to illuminate these points, if the observed effect size was *only* attributed to maturation, we would expect a much smaller posterior mean value of approximately .04, if not even smaller (see analytic plan section for RQ2). Second, we used Bayesian priors informed by previous literature. The research-based priors serve to adjust our posterior in accordance with expectations derived from similar studies, thereby regularizing the posterior mean. In this study, and as seen in figure 1.2, the prior is in fact adjusting our posterior *away* from the likelihood (representing information in the data) towards smaller parameter values. This means that, had we not incorporated informative priors, our posterior mean (or ordinary least squares/maximum likelihood estimate) would have been larger than the reported .18 for NFC.

4.1 Limitations and future directions

Even though our priors circumvent the full force of maturation effects, our results cannot be interpreted as causal. The many weaknesses of a one-group pretest/posttest design are well documented (e.g., Shadish, Cook, & Campbell, 2002); still, these designs are useful in examining implementation, experimenting with new measures, and providing preliminary evidence to support further studies exhibiting more robust study design features. Given the early stage of the module, and the exploratory nature of this initial pilot design, the results from this study can be viewed as preliminary evidence in support of designing a more rigorous IVC evaluation study. Specifically, researchers scaling up the IVC module could evaluate the effects of participation using a randomized control trial where students are randomly assigned to the IVC module within courses. Finally, it would be useful to expand the IVC module to inculcate other intellectual virtues beyond curiosity—such as intellectual humility—and to seek other forms of assessment.

4.2 Conclusion

Higher education is placing increased salience on developing students in a diverse array of skills and dispositions. In line with these developments, the concept of intellectual virtue is beginning to transition from the philosophical literature to empirical assessment in higher education contexts. Concomitantly, with the threat of the COVID-19 virus, higher education is seeing a full transition to online and distance education. The lack of personal connection and traditional opportunities for extra-curricular engagement activities pose challenges to the development of many desirable dispositions. Accordingly, this study was the first, to the authors' knowledge, to introduce and evaluate an online curriculum module developed to inculcate intellectual virtue in university students. In summary, the high level of satisfaction, consistent enjoyment across important student subgroups, and large growth on measures of curiosity and self-reported knowledge and value of intellectual virtue, suggests a robust set of preliminary evidence in support of scaling up the IVC module and conducting further research. Provided the significant changes required for higher education to prioritize habits of mind over success metrics, such as administrative and institutional buy-in (Baehr, 2016), the IVC module may potentially offer educators a feasible, cost-effective means of developing intellectual virtue without relying on a system overhaul.

References

Alves, H., & Raposo, M. (2007). Conceptual model of student satisfaction in higher education. *Total Quality Management*, *18*(5), 571-588.

Armbruster, P., Patel, M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE—Life Sciences Education*, *8*(3), 203-213.

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. OUP Oxford.

Baehr, J. (2013). Educating for intellectual virtues: From theory to practice. *Journal of Philosophy of Education*, *47*(2), 248–262.

Baehr, J. (2016). Is intellectual character growth a realistic educational aim? *Journal of Moral Education*, *45*(2), 117–131.

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, *27*(3), 353–389.

Battaly, H. (2006). Teaching intellectual virtues: Applying virtue epistemology in the classroom. *Teaching Philosophy*, *29*(3), 191–222.

Battaly, H. (2014). *Intellectual virtues*. *The handbook of virtue ethics (ed.) S. van Hooft* (pp. 177–187). London: Acumen.

Byerly, T. R. (2019). Teaching for intellectual virtue in logic and critical thinking classes: Why and how. *Teaching Philosophy*.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

Carter, J. A. (2017). Intellectual autonomy, epistemic dependence and cognitive enhancement. *Synthese*, 1–25.

Castle, T. D. (2014). The impact of cooperative learning on the development of Need for Cognition among first-year college students (Doctoral dissertation).

Day, E. A., Espejo, J., Kowollik, V., Boatman, P. R., & McEntire, L. E. (2007). Modeling the links between need for cognition and the acquisition of a complex skill. *Personality and Individual Differences*, *42*(2), 201–212.

Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: the mediator role of Students' expectancy-value beliefs. *Frontiers in psychology*, *8*, 1193

Dyer, K. D., & Hall, R. E. (2019). Effect of critical thinking education on epistemically unwarranted beliefs in college students. *Research in Higher Education*, *60*(3), 293–314.

Elias, S. M., & Loomis, R. J. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance 1. *Journal of Applied Social Psychology*, *32*(8), 1687–1702.

Eren, A., & Coskun, H. (2016). Students' level of boredom, boredom coping strategies, epistemic curiosity, and graded performance. *The Journal of Educational Research*, *109*(6), 574–588.

Fagioli, L. P., Baker, R., & Orona, G. A. (2020). The role of non-cognitive variables in identifying community college students in need of targeted supports. *Research in Higher Education*, 1-39.

Furnham, A., & Thorne, J. D. (2013). Need for cognition: Its dimensionality and personality and intelligence correlates. *Journal of Individual Differences*, *34*(4), 230.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38.

Heersmink, R. (2018). A virtue epistemology of the internet: Search engines, intellectual virtues and education. *Social Epistemology*, *32*(1), 1–12.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, *41*(2), 111-127.

Hyslop-Margison, E. J. (2003). The failure of critical thinking: Considering virtue epistemology as a pedagogical adventure. *Philosophy of Education Archive*, 319–326.

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177.

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., ... & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, *43*(6), 793-813.

Levy, R. (2016). Advances in Bayesian modeling in educational research. *Educational psychologist*, *51*(3-4), 368-380.

Lin, P. C., & Chen, H. M. (2016). The effects of flipped classroom on learning effectiveness: Using learning satisfaction as the mediator. *World Transactions on Engineering and Technology Education*, *14*(2), 231-244.

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, *44*(7), 1585–1595.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, *80*(1), 75–86.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in r and stan*. CRC press.

McGrath, R. E., Hall-Simmonds, A., & Goldberg, L. R. (2020). Are measures of character and personality distinct? Evidence from observed-score and true-score analyses. *Assessment*, *27*(1), 117-135.

Miller, A. N., & Orsillo, S. M. (2020). The impact of values congruity on underrepresented minority (URM) graduate student psychosocial functioning. Higher Education Research & Development, 1-16.

Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 68–75.

Neigel, A. R., Behairy, S., & Szalma, J. L. (2017). Need for cognition and motivation differentially contribute to student performance. *Journal of Cognitive Education and Psychology*, *16*(2), 144–156.

Orona, G. A. (2021a). Philosophy's rematch: A new conceptualization of the study of higher education. *Arts and Humanities in Higher Education*, 14740222211002967.

Orona, G. A. (2021b). Broken promises? examining the effectiveness of promising practices in STEM lectures by student subgroups. *Innovative Higher Education*, *46*(2), 223-239.

Overbaugh, R. C., & Nickel, C. E. (2011). A comparison of student satisfaction and value of academic community between blended and online sections of a university-level educational foundations course. *The Internet and Higher Education*, *14*(3), 164-174.

Pascarella, E. T., Wang, J.-S., Trolian, T. L., & Blaich, C. (2013). How the instructional and learning environments of liberal arts colleges enhance cognitive development. *Higher Education*, *66*(5), 569–583.

Pritchard, D. H. (2013). Epistemic virtue and the epistemology of education. *Journal of Philosophy of Education* 47, 236-47.

Pritchard, D. H. (2019). Intellectual virtues and the epistemic value of truth. *Synthese*, DOI: 10.1007/s11229-019-02418-z.

Pritchard, D. H., Millar, A., & Haddock, A. (2010). *The nature and value of knowledge: Three investigations*. OUP Oxford.

Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press.

Rogaten, J., Rienties, B., Sharpe, R., Cross, S., Whitelock, D., Lygo-Baker, S., & Littlejohn, A. (2019). Reviewing affective, behavioural and cognitive learning gains in higher education. *Assessment & Evaluation in Higher Education*, *44*(3), 321–337.

Ross, L. (2020). The virtue of curiosity. *Episteme*, *17*(1), 105–120.

Salama-Younes, M. (2016). Need for cognition and academic achievement. *International Journal of Education and Psychological Research*, *5*(3), 100–108.

Santini, F. D. O., Ladeira, W. J., Sampaio, C. H., & da Silva Costa, G. (2017). Student satisfaction in higher education: a meta-analytic study. *Journal of Marketing for Higher Education*, *27*(1), 1-18.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. B. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research*, *4*(3), 248-267.

Schwartz, B. (2020). Science, scholarship, and intellectual virtues: A guide to what higher education should be like. *Journal of Moral Education*, 1-12.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin,

Turri, J., Alfano, M., & Greco, J. (1999). Virtue epistemology. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/entries/epistemology-virtue/

Wang, J.-S. (2013). The effects of deep approaches to learning on students' need for cognition over four years of college.

Williams, L. B., Bourgault, A. B., Valenti, M., Howie, M., & Mathur, S. (2018). Predictors of underrepresented nursing students' school satisfaction, success, and future education intent. Journal of Nursing Education, 57(3), 142-149.

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zmigrod, L., Zmigrod, S., Rentfrow, P. J., & Robbins, T. W. (2019). The psychological roots of intellectual humility: The role of intelligence and cognitive flexibility. *Personality and Individual Differences*, *141*, 200-208

Tables & Figures

Table 1.1
Summary Statistics by Student and Sample Characteristics

	Full Course Sample	Participant Sample	Completer Sample
Demographics[a]	N | M (SD)	N | M (SD)	N | M (SD)
Male	241 | 0.40 (0.49)	98 | 0.37 (0.48)	73 | 0.36 (0.48)
Female	361 | 0.60 (0.49)	166 | 0.63 (0.48)	129 | 0.64 (0.48)
Non-URM	385 | 0.64 (0.48)	169 | 0.64 (0.48)	129 | 0.64 (0.48)
URM	217 | 0.36 (0.48)	95 | 0.36 (0.48)	73 | 0.36 (0.48)
Non-low-income[b]	415 | 0.69 (0.46)	172 | 0.65 (0.48)	129 | 0.64 (0.48)
Low-income[b]	187 | 0.31 (0.46)	92 | 0.35 (0.48)	73 | 0.36 (0.48)
Non-first-generation	325 | 0.54 (0.50)	153 | 0.58 (0.50)	121 | 0.60 (0.49)
First-generation	277 | 0.46 (0.50)	111 | 0.42 (0.50)	81 | 0.40 (0.49)
Non-transfer student[c]	584 | 0.97 (0.18)	256 | 0.97 (0.18)	196 | 0.97 (0.18)
Transfer student[c]	18 | 0.03 (0.18)	8 | 0.03 (0.18)	6 | 0.03 (0.18)
Course Information[a]			
Philosophy1	304 | 0.50 (0.50)	111 | 0.42 (0.49)	79 | 0.39 (0.49)
Philosophy4	253 | 0.42 (0.49)	116 | 0.44 (0.50)	91 | 0.45 (0.50)
Nursing110W	45 | 0.07 (0.26)	37 | 0.14 (0.34)	32 | 0.16 (0.37)
Course Grade[d]	| 3.72 (0.57)	| 3.82 (0.38)	| 3.87 (0.31)
Term Information[d]			
Quarter GPA	| 3.39 (0.58)	| 3.50 (0.49)	| 3.56 (0.44)
Quarter Units Earned	| 14.58 (3.24)	| 14.65 (3.09)	| 14.73 (2.91)
Cumulative GPA	| 3.20 (0.45)	| 3.29 (0.43)	| 3.33 (0.42)
Prior Academic Achievement[d]			
Previous School GPA	| 3.93 (0.41)	| 4.00 (0.34)	| 4.02 (0.23)
SAT: Math	| 626.11 (93.48)	| 630.00 (88.47)	| 634.90 (87.00)
SAT: Read	| 580.60 (91.35)	| 586.97 (90.98)	| 594.21 (91.86)
Major Cluster[a]			
STEM & Health Sciences	298 | 0.50 (0.50)	152 | 0.58 (0.50)	123 | 0.61 (0.49)
Art & Humanities	28 | 0.05 (0.21)	8 | 0.03 (0.17)	4 | 0.02 (0.14)
Social Science	182 | 0.30 (0.46)	70 | 0.27 (0.44)	52 | 0.26 (0.44)
Business	78 | 0.13 (0.34)	29 | 0.11 (0.31)	21 | 0.10 (0.31)
Undeclared	16 | 0.03 (0.16)	5 | 0.02 (0.14)	2 | 0.01 (0.10)
Total	602	264	202

Note. The completer sample is the sample of students who completed both surveys (pre and post). This is the analytic set which is used in the subsequent analysis. The participant sample is the sample of students who started the surveys but did not complete the post survey. The full course sample is the entire number of students in the courses where the IVC module was implemented. [a]The subgroups listed here are dichotomous variables coded as 1; the means reported for these dichotomous variables is the percentage of students in that subgroup out of the total for the corresponding column.

[b]Low-income students are designated administrative labels based on the university system. Here, this definition includes students who qualify for federal aid due to household incomes of $50,000 or less. [c]Transfer students refer to students who transferred from a community college to the university. [d]These variables are continuous; the number of students with data on these variables corresponds to the total listed in the columns. Therefore, the means and standard deviations presented correspond to the totals listed in the columns.

Table 1.2.
Satisfaction with Module by Subgroup: Means and Standard Deviations

Satisfaction (S) Variable	Overall sample	Gender		URM Status		Low-Income Status		First-Generation Status	
		Male	Female	Non-URM	URM	Non-Low-Income	Low-income	Non-First-generation	First-generation
S1: Satisfied with Module	4.46 (0.62)	4.52 (0.63)	4.43 (0.62)	4.46 (0.61)	4.46 (0.65)	4.46 (0.63)	4.47 (0.63)	4.41 (0.64)	4.53 (0.59)
S2: Effectiveness of Module at Introducing IV	4.22 (0.78)	4.21 (0.87)	4.23 (0.73)	4.18 (0.83)	4.29 (0.68)	4.18 (0.81)	4.30 (0.72)	4.16 (0.85)	4.32 (0.67)
S3: Relational Benefit of IV Module	6.22 (0.95)	6.15 (1.02)	6.26 (0.91)	6.23 (0.95)	6.21 (0.96)	6.17 (0.99)	6.32 (0.88)	6.15 (1.02)	6.33 (0.84)
S4: Recommend Module to Other	82.52 (16.78)	82.48 (19.28)	82.55 (15.26)	81.98 (17.23)	83.51 (16.00)	80.85 (18.21)	85.48 (13.52)	81.45 (18.45)	84.14 (13.88)
N	202	73	129	130	72	129	73	121	81

Note. Standard deviations displayed in parenthesis. The four satisfaction variables were measured on different response scales. The first two listed were positioned on a 5-point scale, with 1 = low satisfaction; 5 = high satisfaction. The relational benefit of the IV module was positioned on a 7-point Likert scale with 1 = other extra-credit opportunities provide much more benefit; 7 = this IVC module provides much more benefit than other extra credit opportunities. Finally, the last satisfaction item was positioned on a 100-point scale, with 0 = not at all likely to recommend; 100 = certainly going to recommend.

Table 1.3.
Pre-Post IVC Posterior Summary

	Posterior Mean	Standard Deviation	Lower 95% Credible Interval	Upper 95% Credible Interval	Convergence Diagnostic ($\hat{R}$)
NFC	0.178	0.070	0.064	0.290	1
EC	0.135	0.068	0.024	0.246	1
Importance of IV to Education	0.746	0.089	0.599	0.894	1
Understanding of IV	0.482	0.094	0.324	0.638	1

Note. The auxiliary parameter, σ, is *M* = 1.0 and *SD* = 0 for both models presented. All models were run with four Markov chains for 2,000 iterations each. Because of the one-group pretest/posttest design of this study, we cannot control for other confounding variables. NFC = Need for cognition; EC = Epistemic Curiosity. IV = Intellectual Virtue. Each model includes just one dummy variable indicating pre/post change such that posttest is coded 1 and pretest is coded 0.

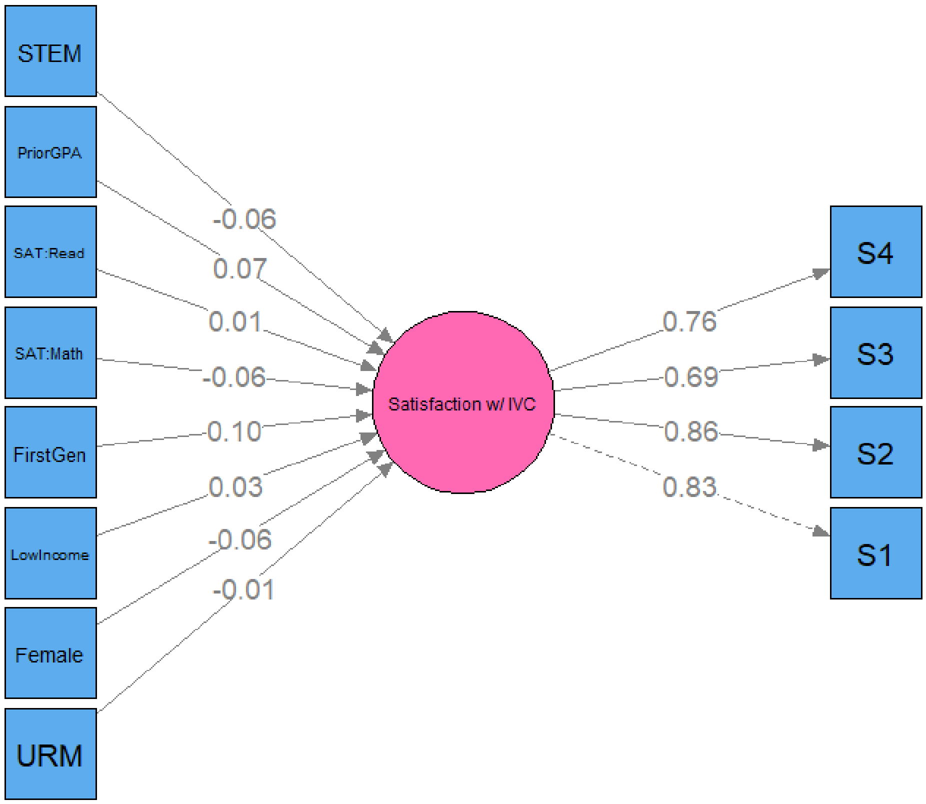


Figure 1.1. Structural equation model displaying a latent satisfaction factor (circle) predicted by student-level variables (left of circle), and four indicators (S1:S4) loading on the satisfaction factor. Loadings and estimates are standardized. None of the regression estimates (left of circle) were statistically significant, $p > .05$; meaning, no significant subgroup differences were found. S1: Satisfied with Module; S2: Effectiveness of Module at Introducing IV; S3: Relational Benefit of IV Module; S4: Recommend Module to Other.

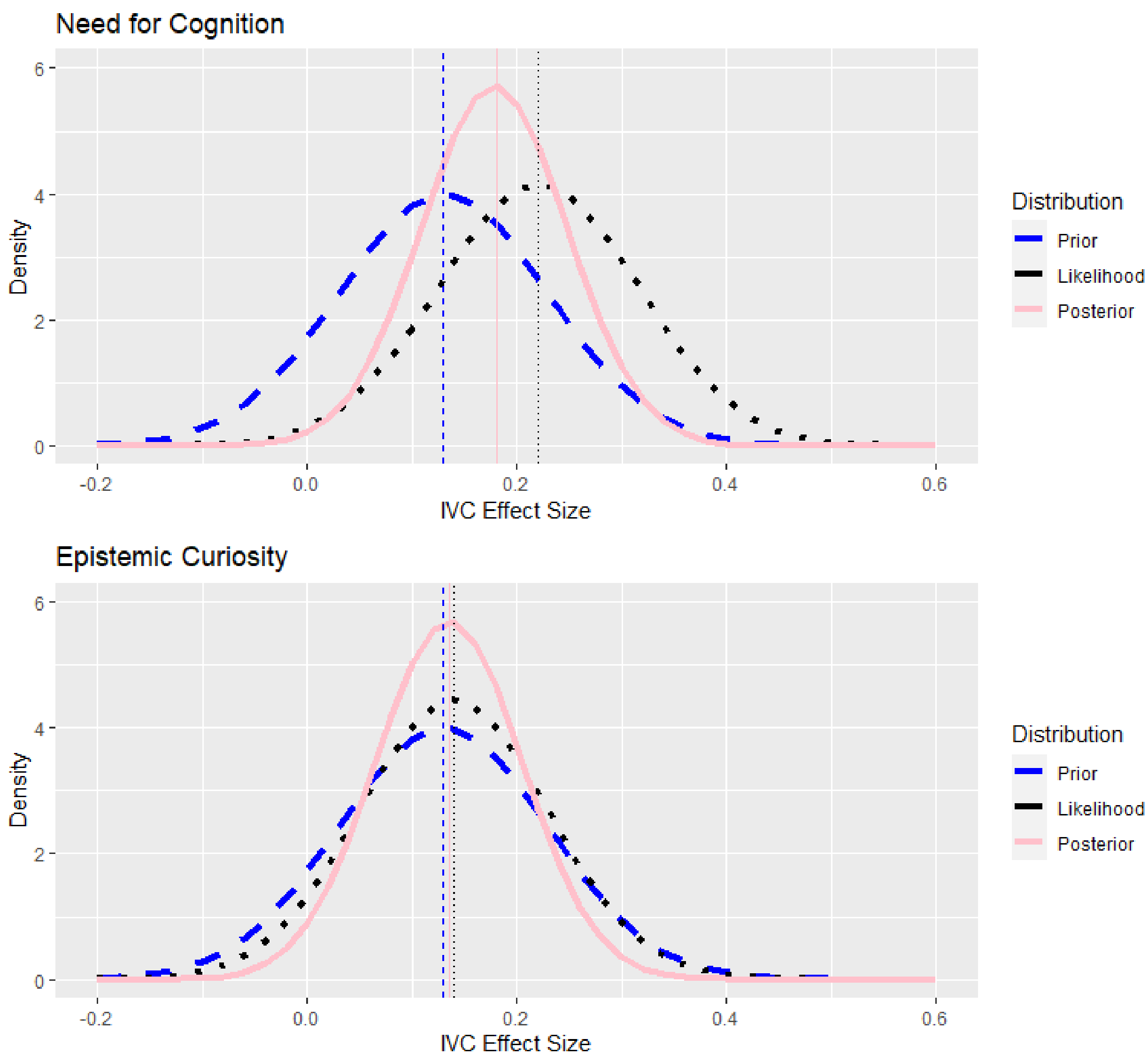


Figure 1.2. Prior, Posterior and Likelihood for the two curiosity scales.

Figure 1.3. Prior and Posterior for the two subjective gain items.

Appendix 1.1: Information and Satisfaction on the IVC Module

Appendix Table 1.1
IVC Applications of Pedagogical Interventions to Develop Interest

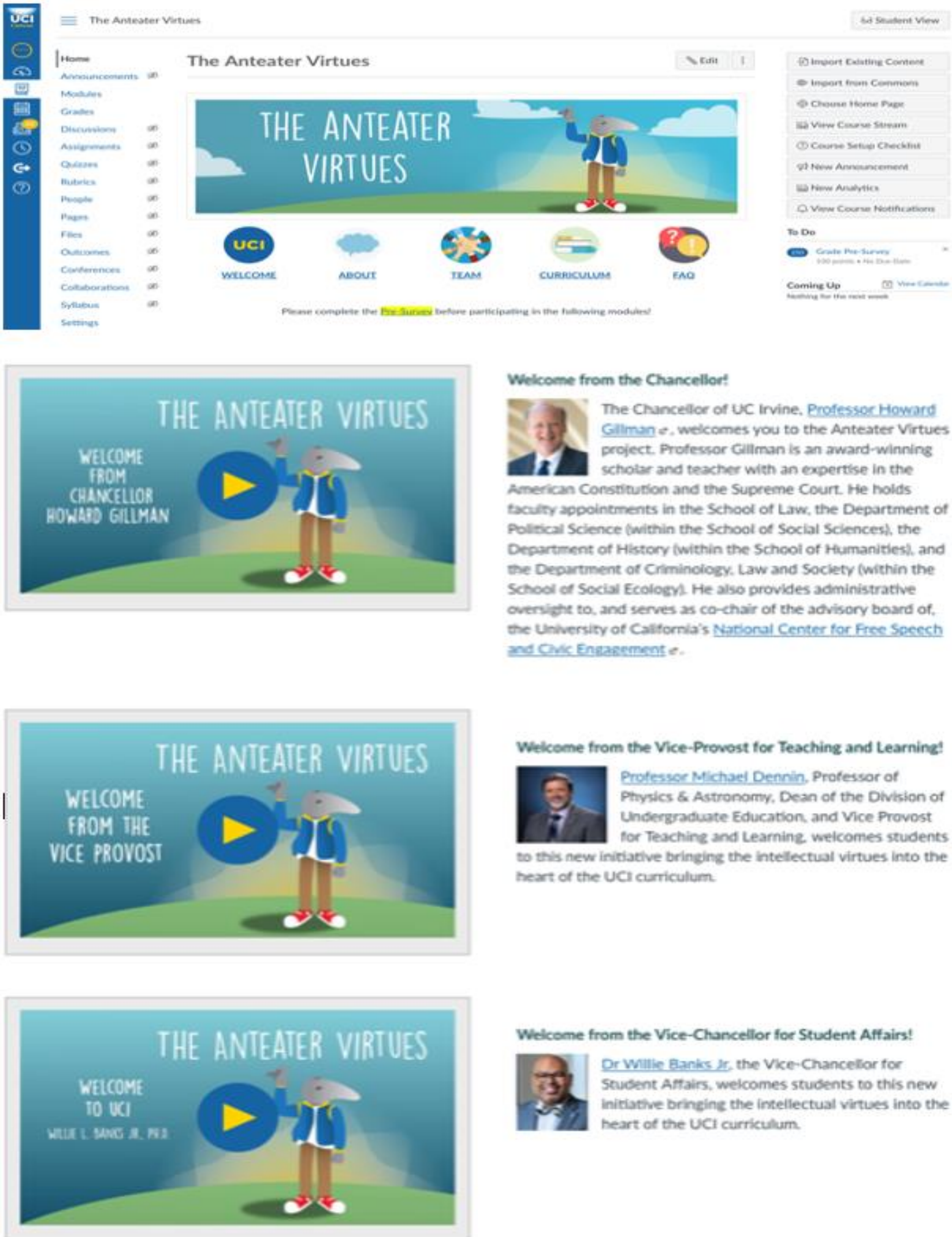
Type of Interest	*Pedagogical Intervention* (and sources)	IVC Application
Triggering situational interest	*Novel experiences* (Palmer, 2009; Maltese & Harsh, 2015; Quinlan, 2019)	Introductory segment that capitalizes on the novelty of the concept of intellectual virtue to students; offers its relevance throughout history as well as for the students' personal educational journey.
	Being exposed to scientists' struggles and applications of concepts (Hong & Lin-Siegler, 2012)	Videos of academics from a range of disciplines describing how a particular intellectual virtue is relevant to their work (e.g., we supply a video from a prominent bioroboticist, filmed in his laboratory, discussing how the intellectual virtue of curiosity is central to his work and necessary to doing quality research).
	Interactive learning activities alongside a lecture course (Yuretich et al., 2001)	Recordings of group activities of fellow university students. The recordings displayed collaborative problem-solving discussions that students were encouraged participate in alongside the video.
Maintaining situational interest	*Repeated involvement (inquiry activities) and novelty (discrepant events)* (Palmer, 2004)	Structured engaging quizzes followed each video. Quizzes were designed such that the process of solving the question(s) reflected intellectual virtue (curiosity) and it's relevance to educational problem-solving.
	Writing activities linking personal goals and values to class content (Hulleman & Harackiewicz, 2009;)	Practical exercises for students to do to reflect on their own development of intellectual virtue, such as recommendations for reading and approaches to educational inquiry.

Note. The table is a modified version of a summary table presented in van der Hoeven Kraft (2017). The types of interest correspond to Hidi & Renniger (2006) phases of interest. In their model, a third phase--supporting individual interest--is presented as the final stage. While research has also applied pedagogical interventions to develop this phase of interest, the one-quarter long, pilot IVC module presented here focuses on the first two phases: triggering situational interest and maintaining situational interest. IVC = Intellectual Virtue Curriculum.

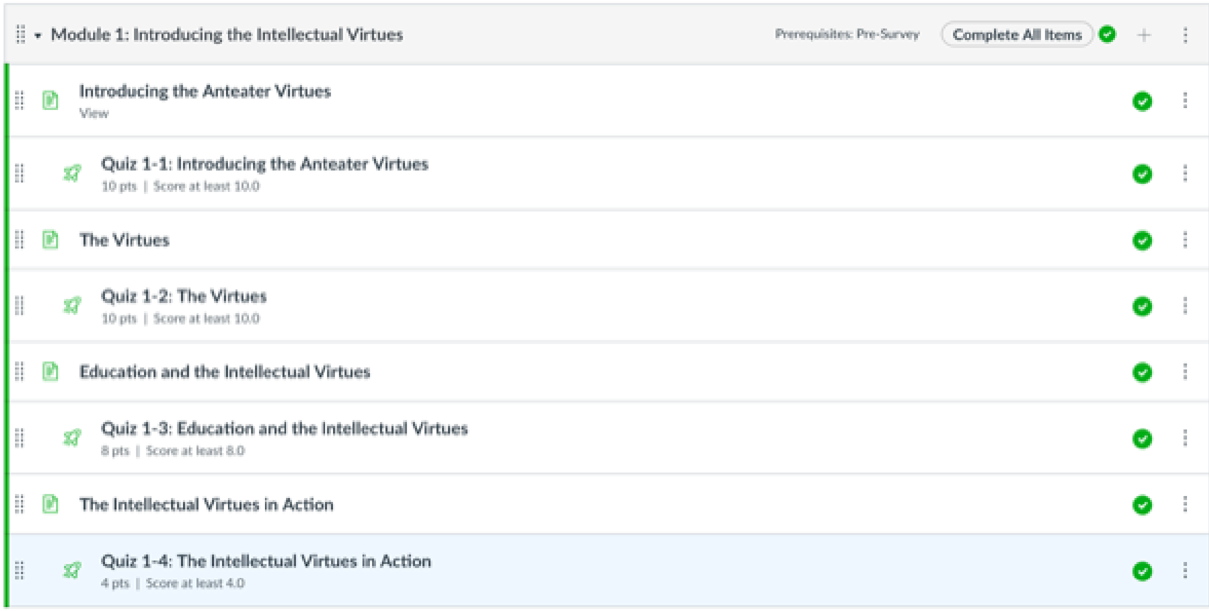
Appendix Table 1.2
Loadings and Regression Coefficients for Satisfaction with Module

	B	SE	*β*	*p*
Factor Loadings				
Satisfied with Module	1.000	0.000	0.825	
Effectiveness of Module at Introducing IV	1.040	0.091	0.859	0.000
Relational Benefit of IV Module	0.831	0.140	0.686	0.000
Recommend Module to Other Students	0.919	0.131	0.758	0.000
Regression Estimates[a]				
URM	-0.015	0.149	-0.009	0.918
Female	-0.105	0.159	-0.061	0.511
Low-income	0.048	0.137	0.028	0.728
First-generation	0.163	0.138	0.097	0.238
SAT: Math	-0.001	0.001	-0.064	0.530
SAT: Read	0.000	0.001	0.013	0.870
Previous School GPA	0.262	0.319	0.074	0.410
Major: STEM	-0.094	0.137	-0.056	0.492

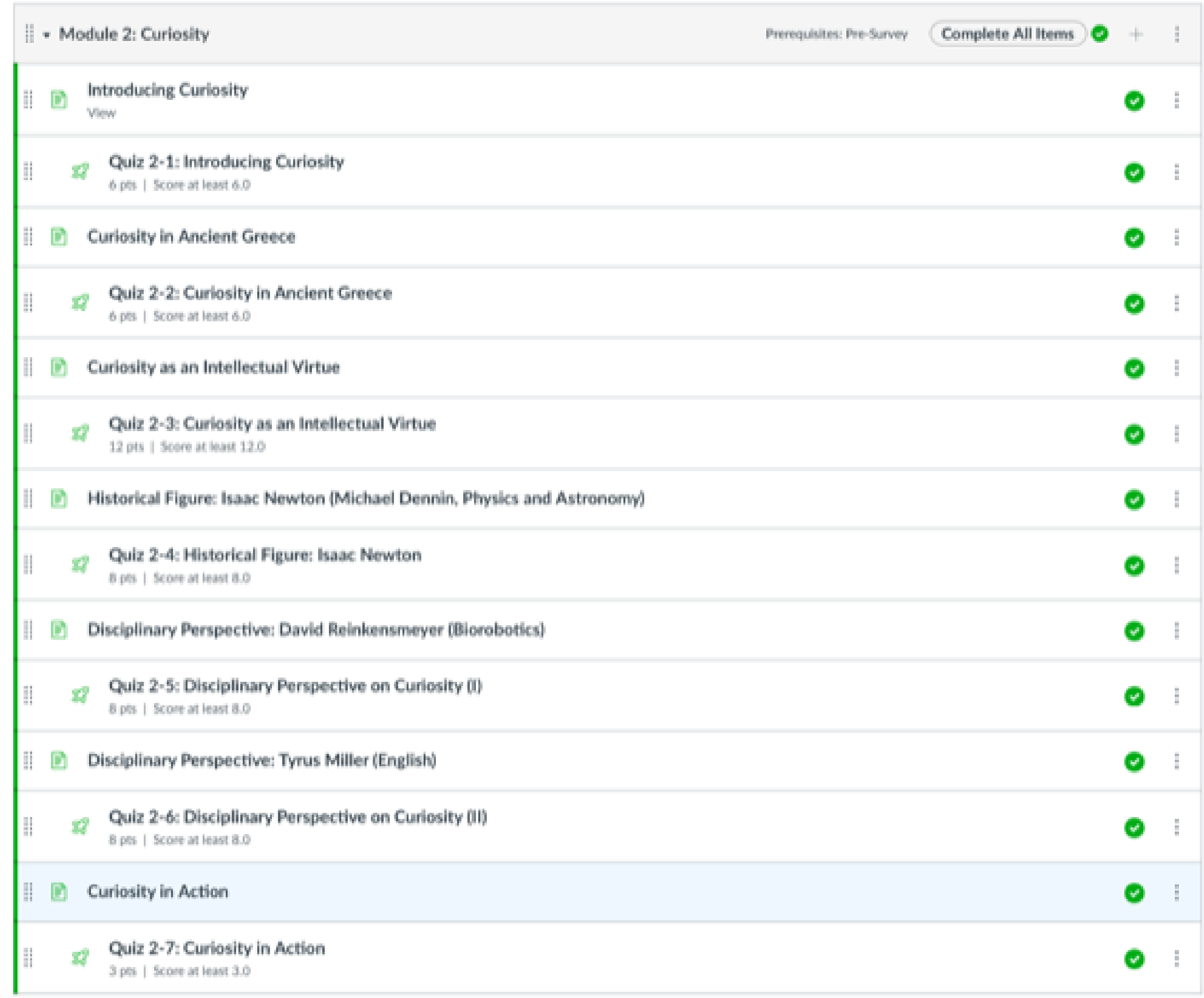
Note. Upper half of table presents factor loadings for latent satisfaction construct. Sine the item, "Satisfied with Module" was designated as the marker variable; *p*-value not displayed for this loading. The lower half of the table presents the regression coefficients where the latent satisfaction construct is regressed on the four demographic predictors presented. Model fit: χ^2 = 49.61 (26), $p < .01$, CFI = .941, TLI = .914, RMSEA = .066, 90% CI (.037, .094). CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Approximation; CI = Confidence Interval.
[a]The subgroups listed here are dichotomous variables coded as 1 for the focal subgroup and 0 for the reference group.



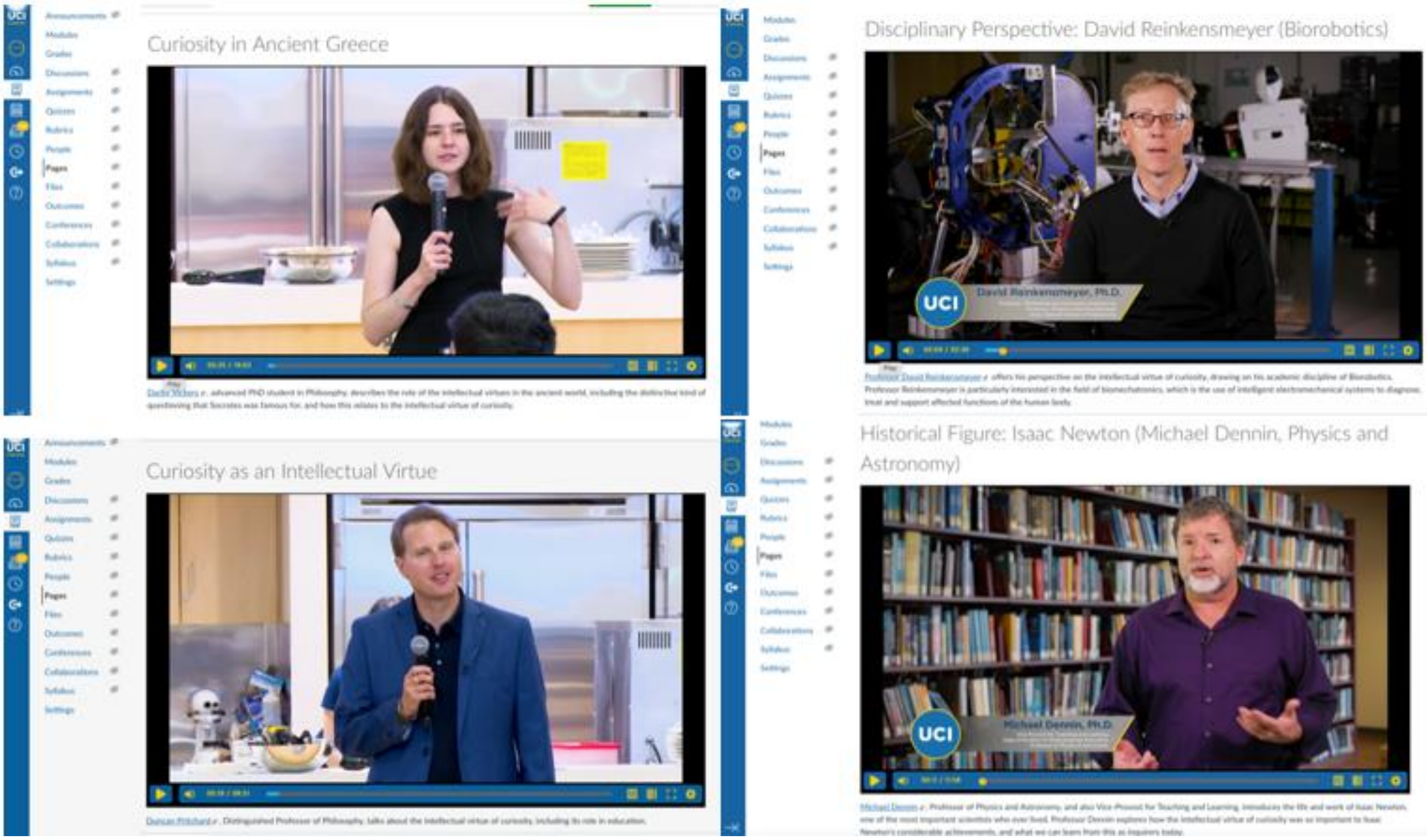
Appendix Figure 1.1: The IVC Welcome Panel



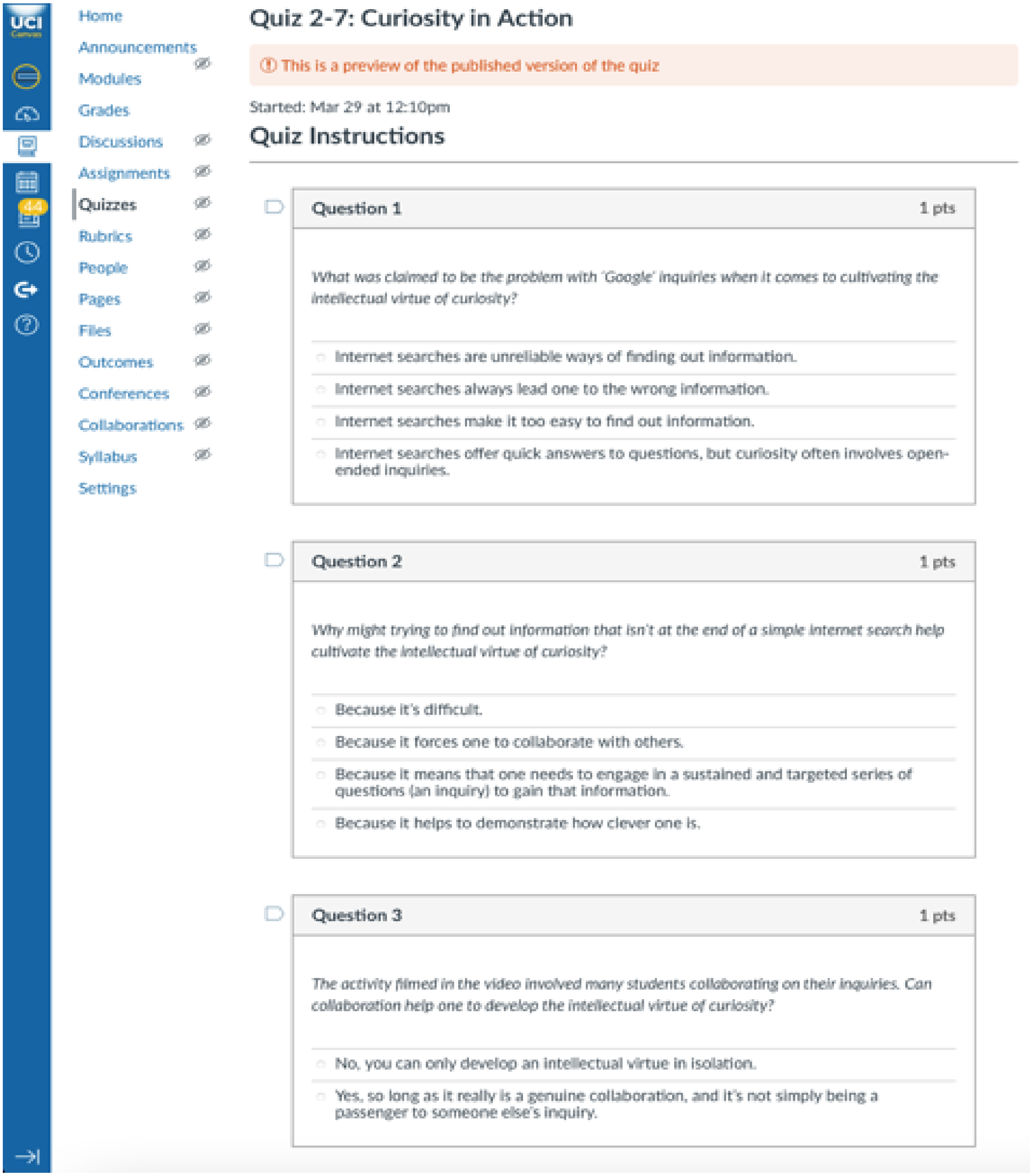
Appendix Figure 1.2: The Introductory Module



Appendix Figure 1.3: The Curiosity Module



Appendix Figure 1.4: Example Lecture Videos of Curiosity Module



Appendix Figure 1.5: Curiosity Quiz Items

CHAPTER 2[3]

Gotta Know Why!

Preliminary Evidence Supporting a Theory of Virtue Learning as Applied to Intellectual Curiosity

1. Introduction

The concept of intellectual virtue is gaining popularity in education (Baehr, 2016; Kotzee, Carter, & Siegel, 2019), and in particular higher education (Hyslop-Margison, 2003; Jones, 2012; Orona, 2021; Schwartz, 2020). The relevance to education has led scholars to supply strategies for developing virtue (e.g., Roberts & Wood, 2007; Battaly, 2016) and explore educational experiences that may cultivate them in students. However, empirical research on the development of intellectual virtue is scant. Moreover, most theoretical stipulations of important relationships remain untested.

In this article, we test a theory of virtue learning as applied to the intellectual virtue of curiosity. First, we introduce Besser's theory which serves in motivating and conceptually framing the present study. Next, we discuss how it relates to intellectual virtue, and subsequently utilize an array of analytic approaches to operationalize and test the hypothesis that increases in the knowledge of (understanding what) and value for (understanding why) intellectual curiosity relates to increases in undergraduate virtue development. This study carries implications for educators interested in pedagogical

[3] This chapter is derived from an article published in *Theory and Research in Education* on 2 Dec 2021, available online: https://doi.org/10.1177/14778785211061310. Full cite: Orona, G. A. (2021). Gotta know why! Preliminary evidence supporting a theory of virtue learning as applied to intellectual curiosity. *Theory and Research in Education*, *19*(3), 279-295.

innovations attempting to enhance students' intellectual virtue in post-secondary education, as well as scholars studying virtue development.

2. Conceptual Framework

Besser (2020) introduces a theory of learning virtue grounded in self-determination theory. Self-determination theory posits that individual perceptions–subjective judgements of the relevance and centrality of actions and goals–are among the foremost drivers of human motivation (Deci & Ryan, 1985, 2012). That is, for goals and habits and activities to be meaningfully adopted in an individual's life, they must resonate with that person's notion of who they are and who they want to be in some sustained way. This kind of intrinsic motivation, in turn, depends on how well decisions satisfy an individual's basic psychological needs, such as autonomy, competence, and relatedness (Deci & Ryan, 1985, 2012; Ryan & Moller, 2017).

Besser (2020) takes the notion of intrinsic motivation as implied by self-determination theory and uses it to construct a theoretical model of virtue development highlighting how the exercise of virtue must resonate with the learner. Besser (2020, p. 287) claims, "…knowing why virtue and its exercise is important places subjects in a position to learn what virtue consists in and how to exercise it…" Resonation is secured through linking the *why* of virtue learning to basic psychological needs, particularly relatedness and autonomy (Besser, 2020).

Accordingly, when learners are reflecting on why virtue resonates with them, they are better situated and equipped to learn *what* virtue consists of and how to exercise it

(Besser, 2020). This, too, reinforces the payoff individuals acquire for practicing virtue, payoff in terms of a sense of relatedness and autonomy. Finally, in addition to the *what* and *why* of virtue learning, the final element in Besser's theory is learning *how*. Knowing how is akin to learning a skill, and once resonation is achieved through reflection on the *what* and *why* of virtue, exercising virtue should become automatic (Besser, 2020).

2.1 Can this be applied to intellectual virtue?

There are at least three key reasons why Besser's theory can be applied to the development of intellectual virtue. First, contemporary virtue epistemologists contend that the intellectual virtues are not especially different than the moral virtues. While the intellectual virtues have been understood in different ways, with some describing them as attitudes or dispositions (Baehr, 2013; Dewey, 1997), others as reliable cognitive skills (Greco, 2000; Sosa, 1985), among the more influential of views suggests a two-component attribute. The two components include "…a motivational component, and a component of reliability in attaining the aims of the motivational component" (Zagzebski, 1996, p. 165).

Thus, the intellectual virtues can be broadly understood as dispositions that are: (a) oriented towards epistemic goods, such as truth and/or knowledge and (b) conducive to acquiring true beliefs (Baehr, 2013, 2015; Pritchard, 2020a; Zagzebski, 1996). In this case, intellectual virtues–just like moral virtues–are character attributes; while one entails good or right actions (moral virtues), the other good or right thinking (intellectual virtues). In fact, Zagzebski (1996) argues that "…no one has offered adequate reason to think the moral and intellectual virtues differ any more than one moral virtue differs from another." (p. 158).

Related to the first point, empirical evidence supports the correspondence between intellectual and moral development. King and Kitchener (1994) note the structural similarities between the conceptual stages of development for reflective and moral judgement (Kitchener, 1982). Reflective judgement, which is a measure of how well individuals' reason about ill-structured problems, has been positively associated with moral reasoning (King, Kitchener, Wood, & Davison, 1989) and different aspects of psycho-social development (Polkosnik & Winston, 1989). Moreover, large-scale studies demonstrate that liberal art experiences result in growth across critical thinking and moral/ethical outcomes (Pascarella, Wang, Trolian, & Blaich, 2013; Seifert, Goodman, King, & Baxter Magolda, 2010). Thus, while Aristotle posited that moral virtue develops through training and habituation and intellectual virtue (aside from being hereditary/bestowed by nature) develops via teaching and instruction, scholars have reformulated Aristotle's original dichotomy of virtue types, emphasizing the blurred and seemingly reciprocal relationship between moral and intellectual development (Battaly, 2016; King & Kitchener, 1994; Perry Jr, 1999; Zagzebski, 1996).

Finally, the third reason why Besser's theory applies to intellectual virtue is implied by the strategies proposed to foster intellectual virtue in students. Proponents of intellectual virtue education highlight the significance of pedagogical techniques such as instruction, the use of exemplars, unique forms of the Socratic method (Watson, 2019), and the opportunity to practice virtuous behavior, as summarized by Kotzee (2019). These strategies correspond to the strategies proposed for moral development (e.g., Besser, 2020; Zagzebski, 1996). Moreover, Battaly (2016) suggests "…formal instruction—lecturing about intellectual virtues and their value—introduces students to new categories, which they can apply to the world and themselves" (p. 173). Here, we see a striking resemblance to the elements in Besser's theory mentioned above: a clear *what* and *why* of virtue learning is considered essential and present on both accounts.

2.2 Current Study

To summarize, the three points show that (a) moral and intellectual virtues are viewed as largely similar (or at least not uniquely different) in kind; (b) development in moral and intellectual domains appear to occur concomitantly, follow an analogous developmental sequence, and are responsive to similar educational arrangements; and (c) the pedagogical strategies recommended for the development of intellectual virtue correspond to the elements in Besser's theory of virtue learning. If the intellectual virtues are indeed a subset of moral virtues with no special distinction between them and other virtues (Zagzebski, 1996), and intellectual curiosity is an intellectual virtue (Ross, 2020), then it follows that a model of virtue development would adequately describe and include intellectual curiosity. In this study, we operationalize Besser's theory of virtue learning in the context of university education. Using an array of analytic techniques, we test the hypothesis that learning the *why* and *what* of intellectual curiosity relates to increases in intellectual curiosity. Thus, the specific contributions of this study include: (i) testing a specific theory as it pertains to undergraduate development of intellectual curiosity and (ii) being among the first empirical studies to examine the adequacy of a general virtue learning framework when applied to intellectual virtue.

3. Methods

3.1 Participants

This study takes place at a large public university located in southern California. There were 202 undergraduate students who participated in this study and who had full data on pretest and posttest measures. As reported by Author (2021), 64% of the sample was female, 36% were underrepresented minorities (URM) and low-income, and 40% were first-generation. 61% of students were declared science, technology, engineering, mathematics and Health Science majors (STEM); the other 39% were spread between social science majors (26%), business (10%), arts and humanities (2%), and undeclared (1%).

3.2 Procedure

These data were collected as part of a larger university-wide project implementing an online curriculum to enhance undergraduate intellectual virtue (Author, 2021). Surveys were administered to students pre and post participation in the online module. It's important to note that this study is not an evaluation of the online module, but rather a correlational test to determine if students who increased on the elements specified in Besser's theory (what and why of virtue learning) display a concomitant increase in intellectual curiosity. For more information on the preliminary findings of the pilot evaluation of the module, please see Orona and Pritchard (2021).

3.3 Measures

Intellectual curiosity. It should be reemphasized that intellectual curiosity is understood as an intellectual virtue. Watson (2018) states its significance: "The identification of curiosity as a basic or fundamental motivating intellectual virtue highlights the special significance of curiosity in an educational setting and, specifically, for intellectual character education". In the empirical sciences, intellectual curiosity (IC) has typically been measured with the following tests: Openness to experience (of the Big-five personality traits), the Typical Intellectual Engagement (TIE) scale, the Epistemic Curiosity (EC) scale, and the Need for Cognition (NFC) scale (Powell, Nettelbeck, & Burns, 2016; Von Stumm, 2013; Von Stumm, Hell, & Chamorro-Premuzic, 2011). Strong correlations have been found between all four, with evidence indicating a unidimensional construct between NFC and EC (Mussel, 2010).

In this study, we collect data on these latter two measures of IC at two time points. Thus, IC is here composed of two instruments: the 18-item Need for Cognition (NFC), and the 5-item Epistemic Curiosity (EC) scale. As these instruments have been validated many times in previous research (e.g., Mussel, 2010; Powell et al., 2016), we combine the two total scores at each time point. Finally, we subtract IC at time 1 from IC at time 2 to obtain the change score.

Figure 2.1 and table 2.1 display the distribution and summary statistics for the IC variables, respectively. As a reminder, the IC variables are sums between the total scores of NFC (measured on a 5-point scale) and EC (measured on a 4-point scale). Therefore, the range of the IC variables is from 1 to 9. All three variables are roughly normally distributed, though the time 2 variable shows some negative skew.

The *What* and *Why*. Operationalizing Besser's *what* and *why*, we utilize data relating to how much students knew about the intellectual virtues (specifically curiosity), and how important they perceived this virtue to their own education. The exact *what* question is: *"Prior to this module, what was your understanding of intellectual virtues?"*. It is positioned on a 3-point response scale. The exact *why* question is: *"How important do you think intellectual virtues are to your education?"* It is positioned on a 6-point response scale. Both items were measured at two time points. Data relating to students' understanding of *how* to implement intellectual virtue was not collected. Table 2.1 displays the summary statistics and bivariate correlations for these variables.

Control variables. Other variables utilized in this study include demographic and academic variables collected from the university system records. Demographic variables include URM status, sex, STEM major, first-generation status, and low-income status–all of which were coded as either 1 (membership in the listed group) or 0. Academic variables include prior (high school or transfer college) GPA, SAT reading score, and SAT math score. Another variable included is an index for students' satisfaction with the intellectual virtue module, composed of four items (Orona & Pritchard, 2021). This is important to include, as we want to eliminate the potential that associations between variables are merely tracking enthusiasm.

3.4 Data analysis

In this study, we take several approaches to data analysis–all of which serve to test our hypothesis, though in different ways. For the first approach, we emphasize a simple descriptive statistic(s): the percentage of students who increased on IC, given that they increased in understanding the *why* and *what* of virtue, respectively. The purpose of this approach is to build an intuition for how essential *what* and *why* are to changes in IC. For instance, if individuals who don't change in *what*/*why* increase in IC at the same rate as those who do, this suggests the possibility of other paths to growing in IC aside from gaining an understanding of the *what* and *why* of virtue.

The second approach focuses on model comparison. Here, we examine Bayes factors and other model fit indices to see which predictor(s) strike a balance between explaining the current data yet avoiding overfitting. For these models, the dependent variable is the IC change score. Additionally, we examine models that include the control variables listed in the measures section.

Finally, for the third approach, we specify cross-lagged latent variable models. For these models, the associations of interest are between the *what* and *why* of learning virtue and intellectual curiosity measured at the two time points. At time point 2, the time 2 correlations hold constant the same variables at time 1. Thus, instead of using change scores and generating composites of IC, we use sumscores (NFC and EC), parcels, and the full 23 item set (18-item NFC and 5-item EC) as indicators of the latent intellectual curiosity factor(s), specifying three different cross-lagged models. The purpose of these models it to examine the stability of the size and significance of the correlations of interest across different constructions of intellectual curiosity.

The purpose in applying these three approaches is to circumvent the possibility that results were produced by chance. Robustness checks are particularly advantageous given current issues with social science replication and provided these data are non-experimental (Freese & Peterson, 2017). Accordingly, we seek coherence between the approaches.

4. Results

4.1 Analytic approach 1: Proportional group comparisons

For the first analytic approach, we dichotomize the *why*, *what*, and IC change scores into two groups, respectively: those who increased on these variables, and those who did not. Table 2.2 showcases the percentage of students who either increased in IC or not (the rows) by whether they increased or not on *what* and *why*. Among those who increased on *why*, 68% also increase in IC. Among those who did not increase on *why*, 61% increased in IC. This means that there is a 7% greater probability of increasing IC if one has also increased in learning why intellectual virtue is important. However, this association was not significant, $\chi^2(1) = 0.77$, $p = 0.38$.

Students who did and did not increase on *what* increased in IC with the same proportion (64%; $\chi^2(1) = 0$, $p = 1$). This means that, descriptively, knowing if someone increased in their knowledge of intellectual virtue is virtually uninformative with respect to knowing if they will also increase their IC. However, categorizing the data in this fashion provides only a broad overview of the relationship; to understand the extent to which *what* and *why* are useful predictors of IC change is more formally examined with the next approach.

4.2 Analytic approach 2: Model comparison

As shown in table 2.3, we specify a variety of models. The models are ranked from the highest Bayes factor to the lowest. The Bayes factor compares two hypotheses (or two models) to one another. All models are compared against the intercept-only model. Values greater than 1 suggest evidence in favor of the specified model in comparison to the intercept-only model (no predictors). Values less than 1 suggest evidence against the shown specified model relative to the intercept-only model (Makowski, Ben-Shachar, & Lüdecke, 2019).

The Bayes factor, Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) offer penalties for models with additional predictors. We see that that the model with only the *why* change score (M1) has the highest Bayes factor, and by a large extent. This model is followed by the model which adds the satisfaction index alongside the why change score (M2). These two models also have the lowest BIC values.

Interestingly, models with the *what* change score do not seem to explain the data well. Additionally, neither do the models with demographic and academic variables, suggesting that neither of these sets of variables were particularly important in understanding students' change in IC. To draw the distinction further, consider M1, with one predictor, which explains 4.1% of the variation in IC change while M10, which has 8 predictors, explains about the same (slightly less, actually) amount.

While not the focus of this strategy, we present the standardized regression estimates for *what* and *why* taken from the full model (M12) in figure 2.2 below. Controlling for the satisfaction index, the *what* change score, and academic and demographic variables, the why change score significantly predicts IC change, (β = .18, p < .05). The *what* change score does not significantly predict IC change, (β = .10, p > .05). All coefficients for all models can be found in the supplementary material.

4.3 Analytic approach 3: Cross-lagged latent variable model

Finally, table 2.4 displays the results of the cross-lagged latent variable models. The information presented includes model fit statistics and the associations of interest–the correlations between *what* and IC and *why* and IC for each time point. The time point two correlations control for time 1 variables. The models differ only in the number of manifest variables used as indicators of the latent IC factor for both time 1 and time 2. For instance, model 1 uses the total scores for NFC and EC as indicators of IC (only two variables loading on the factor), while model 3 uses all of the individual items that constitute the NFC (18-items) and EC (5 items). In contrast, for model 2, we generate 3 composites or parcels

(Little, et al., 2013), all of which have 6 NFC items and at least 1 EC item (because there are 5 EC items, one parcel is composed of 1 EC while the others have 2).

Models 1 and 2 fit the data well across the comparative fit index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). Model three approximates fit for the RMSEA measure but is below recommended thresholds (.8) for CFI and TLI. The point of these models, however, is to examine the stability of the correlations of interest. We see that for each model, and for both *what* and *why*, the size and significance of the associations are consistent, though the *what* correlations vary across time points. For instance, at time point 1, the *what* correlations are each significantly associated with IC, with correlations ranging from .20 to .21, $p < .05$. At time point 2, the *what* correlations have a mere 0.014 range across models and are consistently insignificant across models.

The *why* correlations exhibit moderate correlations for all models, with a mere 0.066 and 0.033 correlation range across models for each time point, and are consistently significantly related to IC, irrespective of the manner in which we construct indicators. Consistent with the other two analytic approaches, learning *why* significantly and positively associates with IC across both time points and all model specifications, $p < .05$. Figure 2.3 depicts the model with the total test scores as indicators (M1), showcasing all specified paths.

5. Discussion

In this study, we applied Besser's theory of virtue learning to undergraduate development of intellectual virtue. Culling information from the three analytic approaches, we see in these data that learning *why* seems to be moderately related to IC growth, while *what* does not. These results therefore provide partial–and preliminary–support for Besser's theory. These results have implications for the designing of (intellectual) virtue learning activities/education, the conceptualization of intellectual virtues, and future research in measuring virtue.

As noted by theorists and philosophers, reflection is central to the development of intellectual virtue (Pritchard, 2020b). Reflective activities have been examined in other virtue studies situated in higher education contexts. For instance, the Global Leadership Initiative, presented and evaluated by Brant, et al. (2020), describe the use of seven pedagogical strategies for emerging adults to learn moral virtue, one of which is "reflection on personal experiences" as it pertains to virtuous activity. Using both quantitative and qualitative analysis, Brant et al. (2020) found that many students report that reflective activities are an important aspect in their individual development.

Additionally, Author (2021) present the reflective activities embedded in the intellectual virtue module in which students in this study participated. They report the positive association of participation in the module on both the *what* and *why* of virtue learning (using the label, "subjective gain"), with larger associations found for *why*. The culmination of these studies, alongside the current one, suggest that–based on traditional notions of mediation requirements (e.g., Baron & Kenny, 1986)–learning *why* virtue is important may describe the mechanism by which reflective activities influence intellectual virtue development. Future research is needed to formally test these links, and further research is needed on the optimal levels and types of activities that meaningfully influence one's understanding of the importance of virtue, and to what extent this satisfies basic psychological needs.

Aside from highlighting the significance of reflection and learning *why* for educational purposes, a broader theoretical implication of this study is that the processes of moral and intellectual character formation appear to respond to similar stimuli. While there was no manipulation of an independent variable in this study, nor do we claim any causal relations, we do highlight the compatibility of this data with the theoretical and empirical work that has surfaced thus far (e.g., Watson, 2019; Besser, 2020; Brant et al., 2020), which suggests that the intellectual and moral virtues share pedagogical developmental techniques. This makes sense, if indeed the intellectual virtues are a subset of the moral virtues (Zagzebski, 1996).

This support is particularly interesting in light of recent arguments against pursuing moral education in higher education settings (e.g., Carr, 2017). But, given the above stated connection, how can universities inculcate intellectual virtue without developing their moral analogue (and vice versa)? This does of course hinge upon how the intellectual virtues are being conceptualized, as noted in the beginning of this paper. If educators and researchers move toward a conception of intellectual virtue that articulates a greater influence from cognitive abilities–such as critical thinking and/or reflective judgement–then the connection between moral and intellectual virtues and the shared theories and pedagogies may dissipate or weaken substantially (Kotzee et al., 2019; Siegel, 2015). How useful the intellectual analogues to the moral virtues are for understanding intellectual development in higher education requires rigorous empirical research; future work should focus on how well measures of intellectual virtue relate to other meaningful behaviors, skills, and abilities.

5.1 Limitations and future directions

There are several limitations with the current study. First, this study is correlational. There are no causal links tested, and confidently identifying exogenous variation in the independent variables of interest precludes the current study design. Still, across the various analytic approaches employed, we find consistent results for the two associations of interest. Thus, the results are not contingent on how we operationalized the variables nor what statistical approach was used. This is especially important given current concerns over selective reporting and researcher degree freedom. So while this study is not causal, the correlations are consistent and stable.

Second, while these data include students from different majors and two different classes enrolled at a diverse public institution, greater generalizability can be enhanced by collecting data from multiple institutions. A multi-site study would be advantageous in testing the stability of these associations across settings where individuals and institutional policies and programs vastly differ.

Finally, the third limitation involves measuring the concepts in Besser's theory, and virtue more broadly. For the *what* and *why* operationalized here, we relied on one self-reported item each. How well these items capture the intended construct(s) is debatable. Performance assessments and other creative data collection procedures can be used to better represent these constructs in future studies.

As for measuring virtue, the use of self-report to measure intellectual curiosity is undoubtedly problematic (Maul, 2017b, 2017a; Ng & Tay, 2020). An approach capturing optimal levels of behavior can be used to develop a more precise measure of intellectual curiosity (Ng & Tay, 2020), which can be subsequently tested for its utility in predicting accurate and calibrated epistemic beliefs. While we acknowledge this weakness, at this early stage of intellectual virtue theory testing, we regard the reliance on the EC and NFC–the latter enjoying a persistent 40-year research base (e.g., Cacioppo & Petty, 1982; Lavrijsen, et al., 2021)–as a reasonable starting point.

5.2 Conclusion

In this study, we argued for the applicability of a general virtue learning theory in describing intellectual virtue development. We then tested two of three critical elements (*what* and *why*, but not *how*, of virtue learning) as they relate to increases in intellectual curiosity measured at two time points. We find consistent yet moderate correlations between intellectual curiosity and learning *why* virtue is important. Weaker associations are found with increasing knowledge of intellectual virtue. These results suggest partial support for the elements of Besser's theory as they apply to intellectual virtue. The implications of these results connect with pedagogical recommendations stressed across intellectual and moral virtue development education.

References

Arum R., Eccles J.S., Heckhausen J., et al. (2021) Ensuring a More Equitable Future: Assessing Student Learning and Growth in Higher Education. *Postsecondary Value Commission*.

Baehr, J. (2013). Educating for intellectual virtues: From theory to practice. *Journal of Philosophy of Education*, *47*(2), 248–262.

Baehr, J. (2015). The four dimensions of an intellectual virtue. In Chienkuo Mi Michael Slote & E. Sosa (Eds.), *Moral and intellectual virtues in western and chinese philosophy* (pp. 86–98.). Routledge.

Baehr, J. (2016). Is intellectual character growth a realistic educational aim? *Journal of Moral Education*, *45*(2), 117–131.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Battaly, H. (2016). Responsibilist virtues in reliabilist classrooms. *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, 163–187.

Besser, L. L. (2020). Learning virtue. *Journal of Moral Education*, *49*(3), 282–294.

Brant, J., Lamb, M., Burdett, E., & Brooks, E. (2020). Cultivating virtue in postgraduates: An empirical study of the oxford global leadership initiative. *Journal of Moral Education*, *49*(4), 415–435.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.

Carr, D. (2017). Virtue and character in higher education. *British Journal of Educational Studies*, *65*(1), 109–124.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory.

Dewey, J. (1997). *How we think*. Courier Corporation.

Freese, J., & Peterson, D. (2017). Replication in social science. *Annual Review of Sociology*, *43*, 147–165.

Greco, J. (2000). Two kinds of intellectual virtue. *Philosophy and Phenomenological Research*, *60*(1), 179–184.

Hyslop-Margison, E. J. (2003). The failure of critical thinking: Considering virtue epistemology as a pedagogical adventure. *Philosophy of Education Archive*, 319–326.

Jones, W. E. (2012). Higher education, academic communities, and the intellectual virtues. *Educational Theory*, *62*(6), 695–711.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults. Jossey-bass higher and adult education series and jossey-bass social and behavioral science series.* ERIC.

King, P. M., Kitchener, K. S., Wood, P. K., & Davison, M. L. (1989). Relationships across developmental domains: A longitudinal study of intellectual, moral, and ego development. *Adult Development: Vol. 1. Comparisons and Applications of Adolescent and Adult Developmental Models*, 57–72.

Kitchener, K. S. (1982). Human development and the college campus: Sequences and tasks. *New Directions for Student Services*, *1982*(20), 17–45.

Kotzee, B., Carter, J. A., & Siegel, H. (2019). Educating for intellectual virtue: A critique from action guidance. *Episteme*, 1–23.

Lavrijsen, J., Preckel, F., Verachtert, P., Vansteenkiste, M., & Verschueren, K. (2021). Are motivational benefits of adequately challenging schoolwork related to students' need for cognition, cognitive ability, or both? *Personality and Individual Differences*, *171*, 110558.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285.

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). BayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Maul, A. (2017a). Moving beyond traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 103–109.

Maul, A. (2017b). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 51–69.

Mussel, P. (2010). Epistemic curiosity and related constructs: Lacking evidence of discriminant validity. *Personality and Individual Differences*, *49*(5), 506–510.

Ng, V., & Tay, L. (2020). Lost in translation: The construct representation of character virtues. *Perspectives on Psychological Science*, *15*(2), 309–326.

Orona, G.A. (2021a). Philosophy's rematch: A new conceptualization of the study of higher education. *Arts and Humanities in Higher Education*, 14740222211002967.

Orona, G.A. (2021b). Broken promises? examining the effectiveness of promising practices in STEM lectures by student subgroups. *Innovative Higher Education*, *46*(2), 223-239.

Orona, G.A. and Pritchard, D. (2021). Inculcating curiosity: pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment and Evaluation in Higher Education*, 1-15.

Pascarella, E. T., Wang, J.-S., Trolian, T. L., & Blaich, C. (2013). How the instructional and learning environments of liberal arts colleges enhance cognitive development. *Higher Education*, *66*(5), 569–583.

Perry Jr, W. G. (1999). *Forms of intellectual and ethical development in the college years: A scheme. Jossey-bass higher and adult education series.* ERIC.

Polkosnik, M. C., & Winston, R. B. (1989). Relationships between students' intellectual and psychosocial development: An exploratory investigation. *Journal of College Student Development*.

Powell, C., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences*, *95*, 147–151.

Pritchard, D. (2020a). Educating for intellectual humility and conviction. *Journal of Philosophy of Education*, *54*(2), 398–409.

Pritchard, D. (2020b). Knowledge, skill and virtue epistemology. In *The routledge handbook of philosophy of skill and expertise* (pp. 135–145). Routledge, Taylor; Francis Group.

Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press.

Ross, L. (2020). The virtue of curiosity. *Episteme*, *17*(1), 105–120.

Ryan, R. M., & Moller, A. C. (2017). Competence as central, but not sufficient, for high-quality motivation. *Handbook of Competence and Motivation: Theory and Application*, 216–238.

Schwartz, B. (2020). Science, scholarship, and intellectual virtues: A guide to what higher education should be like. *Journal of Moral Education*, 1–12.

Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. B. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research*, *4*(3), 248–267.

Siegel, H. (2015). Critical thinking and the intellectual virtues. In *Intellectual virtues and education: Essays in applied virtue epistemology* (pp. 95–112). Taylor; Francis Inc.

Sosa, E. (1985). Knowledge and intellectual virtue. *The Monist*, *68*(2), 226–245.

Von Stumm, S. (2013). Investment traits and intelligence in adulthood. *Journal of Individual Differences*.

Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574–588.

Watson, L. (2018). Educating for curiosity. In I. Inan, L. Watson, D. Whitcomb, & S. Yigit (Eds.), *The moral psychology of curiosity* (pp. 293–309). Rowman & Littlefield International.

Watson, L. (2019). Educating for inquisitiveness: A case against exemplarism for intellectual character education. *Journal of Moral Education*, *48*(3), 303–315.

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Tables and Figures

Table 2.1
Means, Standard deviations, and Correlations.

				Time 1 Variables					Time 2 Variables						Change Score (Difference Variables)	
	M	SD	α	NFC_1	EC_1	IC_1	Why_1	What_1	NFC_2	EC_2	IC_2	Why_2	What_2	Satisfaction	IC Change	Why Change
NFC_1	3.44	0.61	.87													
EC_1	3.14	0.6	.88	0.47***												
IC_1	6.58	1.04	.64	0.86***	0.85***											
Why_1	4.38	1.38		0.28***	0.14*	0.24***										
What_1	1.61	0.59		0.21**	0.14*	0.20**	0.19**									
NFC_2	3.57	0.57	.85	0.78***	0.39***	0.69***	0.16*	0.15*								
EC_2	3.23	0.63	.91	0.43***	0.74***	0.68***	0.11	0.12	0.53***							
IC_2	6.8	1.05	.69	0.68***	0.66***	0.78***	0.15*	0.15*	0.86***	0.89***						
Why_2	5.28	0.83		0.38***	0.24***	0.36***	0.28***	0.06	0.42***	0.34***	0.44***					
What_2	1.93	0.76		0.15*	0.02	0.1	0.09	0.46***	0.12	0.12	0.14*	0.13				
Satisfaction	0	1	.86	0.26***	0.25***	0.29***	0.23***	0.01	0.33***	0.34***	0.39***	0.50***	0			
IC Change	0.22	0.69		-0.26***	-0.29***	-0.32***	-0.14	-0.07	0.27***	0.32***	0.34***	0.12	0.06	0.14*		
Why Change	0.90	1.4		-0.05	0	-0.03	-0.82***	-0.15*	0.09	0.09	0.11	0.32***	-0.01	0.07	0.20**	
What Change	0.32	0.71		-0.01	-0.09	-0.06	-0.06	-0.34***	0.01	0.03	0.02	0.08	0.68***	-0.01	0.12	0.11

Note. M = mean; SD = standard deviation. α= Cronbach's alpha. _1 = time 1; _2 = time 2 NFC = need for cognition; EC = epistemic curiosity; IC = Intellectual curiosity.

Table 2.2
Group Proportions

	Why		*What*		Total
IC	No Increase	Increased	No Increase	Increased	
No Increase	37 (39%)	35 (32%)	52 (36%)	20 (36%)	72 (36%)
Increased	57 (61%)	73 (68%)	94 (64%)	36 (64%)	130 (64%)
Colum Total	94	108	146	56	202

Note. IC = intellectual curiosity; *why = How important do you think intellectual virtues are to your education?; what = Prior to this module, what was your understanding of intellectual virtues?*

Table 2.3
Model Comparisons (Ranked by Bayes Factor)

Model	Bayes Factor	R^2	*df*	AIC	BIC
M1: Why Change Score	4.9939	0.041	1	420.306	430.23
M2: Satisfaction + Why Change Score	2.1428	0.058	2	418.69	431.923
M3: Why Change Score + What Change Score	0.9872	0.051	2	420.239	433.473
M4: Satisfaction	0.5775	0.021	1	424.62	434.545
M5: Satisfaction + Why Change Score + What Change Score	0.4498	0.068	3	418.503	435.045
M6: What Change Score	0.3117	0.015	1	425.854	435.778
M7: Satisfaction + What Change Score	0.19059	0.035	2	423.529	436.762
M8: Satisfaction + Why Change Score + Demographics/Academic Variables	6.72E-08	0.094	10	426.78	466.479
M9: Why Change Score + What Change Score + Demographics/Academic Variables	3.92E-08	0.090	10	427.856	467.556
M10: Demographics/Academic Variables	3.19E-08	0.039	8	434.885	467.968
M11: Satisfaction + Demographics/Academic Variables	1.75E-08	0.058	9	432.78	469.171
M12: Satisfaction + Why Change Score + What Change Score + Demographics/Academic Variables	1.54E-08	0.105	1	425.854	435.778
M13: Satisfaction + What Change Score + Demographics/Academic Variables	6.63E-09	0.073	10	431.411	471.111

Note. Demographics/Academic Variables = URM + Female + STEM + First-generation + Low-income + Prior GPA + SAT: Read + SAT: Math. *df* = degrees of freedom; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Table 2.4
Cross-lagged Latent Variable Models

	Model Fit						Time 1 correlations with (standardized) Latent IC Variable		Time 2 correlations with (standardized) Latent IC Variable	
Model	χ2	*df*	CFI	TLI	RMSEA	RMSEA 90% CI	*what*	*why*	*what*	*why*
M1: Sumscores as IC indicators	13.282	8	0.99	0.964	0.059	0-0.113	0.196**	0.286***	0.085	0.332***
M2: 3 Parcels as IC indicators	30.205	22	0.992	0.983	0.046	0-0.083	0.199*	0.201**	0.083	0.313***
M3: All 23 items as IC indicators	2620.8	1140	0.676	0.652	0.085	0.081-0.09	0.208*	0.220*	0.071	0.299**

Note. ** = $p < .01$; *** = $p < .001$. M1 = model 1; M2 = model 2; M3 = model 3. IC = Intellectual curiosity. "*What*" refers to the knowledge of intellectual virtues at time 2; "*why*" refers to the perceived importance of intellectual virtues to student's education at time 2. Parcels are aggregated composites; here, we use 3 parcels, each of which have 6 NFC items and at least 1 EC item (because there are 5 EC items, one parcel is composed of 1 EC while the others have 2). χ2 = chi-square; *df* = degrees of freedom; CFI = comparative fit index, TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation.

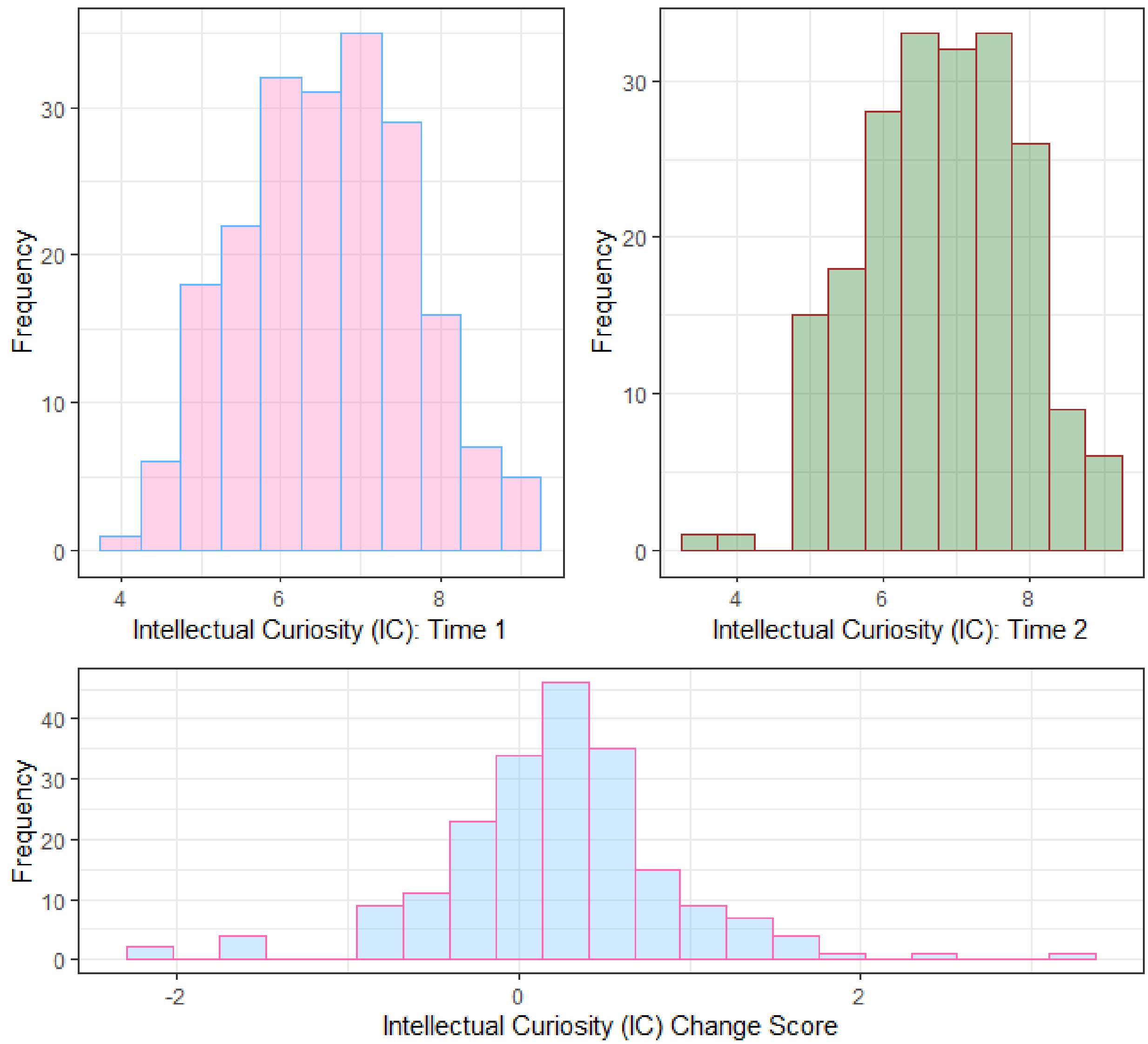


Figure 2.1. Top pane: Intellectual curiosity (IC) at times 1 and 2. Subtracting time 1 IC from time 2 IC generates the IC change score, shown on the bottom pane.

Figure 2.2. Linear relationships for the two elements in Besser's theory with intellectual curiosity (IC). All variables presented are change scores from time 1 and time 2. β estimates are presented from the full model with all controls included (M12 in table 3). * = $p < .05$.

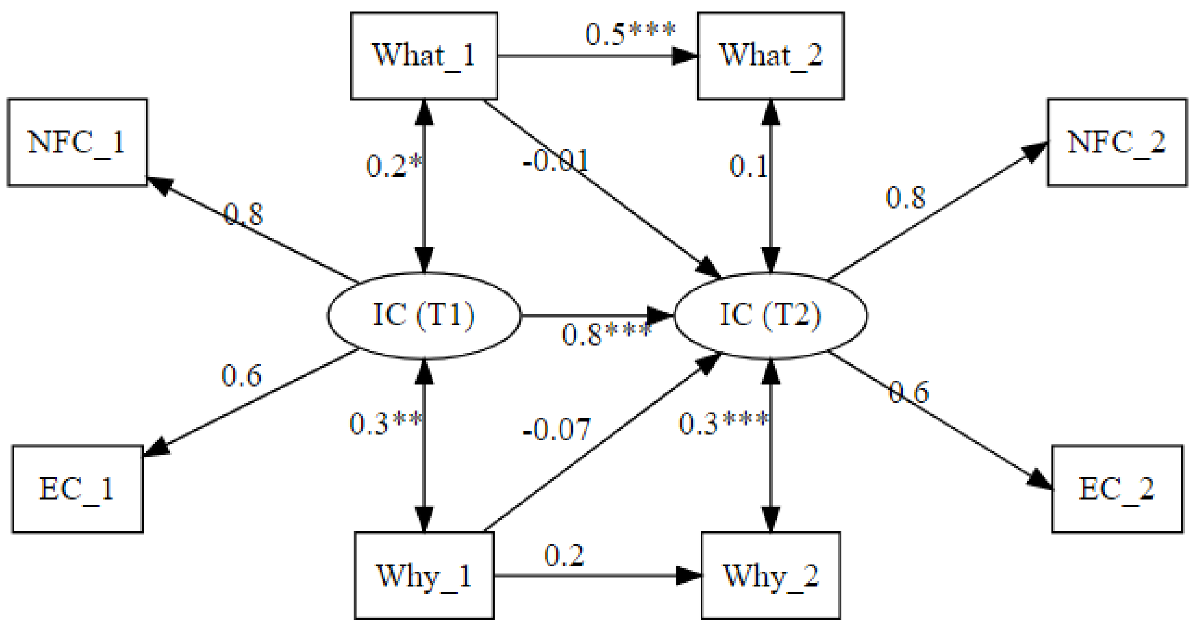


Figure 2.3. Structural equation cross-lagged latent variable model with Need for Cognition (NFC) and Epistemic Curiosity (EC) total test scores as indicators of intellectual curiosity (IC).

Appendix 2.1: Regression Estimates

Appendix Table 2.1
Unstandardized Regression Estimates for Models Presented in Table 2.3

	Dependent variable: IC Change Score (unstandardized coefficients are shown)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
URM	0.090				0.064					0.092	0.071	0.086	0.067
	(0.118)				(0.116)					(0.117)	(0.116)	(0.117)	(0.115)
female	0.125				0.112					0.134	0.127	0.125	0.120
	(0.110)				(0.108)					(0.109)	(0.108)	(0.109)	(0.107)
STEM	-0.119				-0.117					-0.108	-0.117	-0.097	-0.107
	(0.109)				(0.107)					(0.109)	(0.107)	(0.108)	(0.107)
FirstGeneration	0.087				0.083					0.070	0.045	0.097	0.069
	(0.118)				(0.117)					(0.117)	(0.116)	(0.118)	(0.116)
LowIncome	0.037				0.077					0.029	0.067	0.035	0.069
	(0.116)				(0.115)					(0.116)	(0.114)	(0.115)	(0.114)
PrevSchoolGPA	0.146				0.179					0.120	0.148	0.129	0.154
	(0.245)				(0.240)					(0.243)	(0.239)	(0.242)	(0.239)
SAT Read	0.001				0.001					0.001	0.001	0.001	0.001
	(0.001)				(0.001)					(0.001)	(0.001)	(0.001)	(0.001)
SAT Math	0.001				0.001					0.001	0.001	0.001	0.001
	(0.001)				(0.001)					(0.001)	(0.001)	(0.001)	(0.001)
Why Change		0.101***		0.095***	0.095***		0.096***		0.091***		0.096***		0.090**
		(0.034)		(0.034)	(0.035)		(0.034)		(0.034)		(0.035)		(0.035)
What Change			0.117*	0.096	0.101			0.118*	0.098			0.123*	0.102
			(0.068)	(0.067)	(0.068)			(0.067)	(0.067)			(0.068)	(0.068)
Satisfaction Index						0.099**	0.090*	0.100**	0.091*	0.097**	0.087*	0.096**	0.087*
						(0.048)	(0.048)	(0.048)	(0.047)	(0.049)	(0.048)	(0.049)	(0.048)
Constant	-1.552*	0.128**	0.181***	0.102*	-1.679*	0.219***	0.132**	0.181***	0.106*	-1.474	-1.529*	-1.573*	-1.608*
	(0.933)	(0.057)	(0.053)	(0.060)	(0.915)	(0.048)	(0.056)	(0.053)	(0.059)	(0.927)	(0.911)	(0.923)	(0.910)

F Statistic	0.968 (df = 8; 193)	8.621*** (df = 1; 200)	2.969* (df = 1; 200)	5.356*** (df = 2; 199)	1.881** (df = 10; 191)	4.212** (df = 1; 200)	6.164*** (df = 2; 199)	3.662** (df = 2; 199)	4.851*** (df = 3; 198)	1.311 (df = 9; 192)	1.993** (df = 10; 191)	1.515 (df = 10; 191)	2.027** (df = 11; 190)

Note: p<0.1; **p<0.05; ***p<0.01. All models include 202 observations.

CHAPTER 3

Cognitive Development Among Undergraduate Emerging Adults: How Course-Taking Breadth Supports Skill Formation

1. Introduction

Cognitive skills, such as critical thinking, information processing, and reflective judgement, are longstanding aims of higher learning and the liberal arts tradition and remain among the most sought-after competencies in the modern workplace (Arum & Stevens, 2020; Rios, Ling, Pugh, Becker, & Bacall, 2020; Vista, 2020; Wolff & Booth, 2017). While these concepts have been explored for many decades in university settings, previous research has usually relied on cross-sectional data, self-report instruments, single assessments, low-samples, or a mix of all these attributes. Additionally, implications following large-scale impact studies do little to inform our theoretical understanding of the developmental and environmental processes underlying growth among emerging adults.

The purpose of this study is to examine the relationship of course-taking breadth on the formation of complex reasoning skills, situating the current research within the adult cognitive development literature. First, we introduce Fischer's (1980) dynamic skill theory–an enduring integrative cognitive model (Mascolo, 2020), which serves to conceptually frame the present study. Next, we outline the appropriateness of this framework as it pertains to contemporary research in higher education and emerging adult intellectual development. Finally, we provide an analysis of longitudinal data spanning two years, and subsequently discuss the implications for scholars studying adult cognition and student learning.

2. Conceptual Framework

Fischer (1980a) introduced dynamic skill theory to predict and describe the sequence and synchrony of cognitive skills across the lifespan. The primary element of the theory is the skill–which is understood as a person's operant control within a specific context or environment. Sources of variation that individual's control (e.g., acting and thinking) are represented as collections or sets or classes. The combination or coordination of sets of sources of variation comprise skills. Fischer (1980, p. 482) notes, "What makes a group of sets into a skill is the person's control over both each individual set and the relations between the sets. For example, an infant who can shake a rattle in order to listen to it has a skill composed of two related sets, shaking the rattle and listening to the noise it makes."

Eschewing strict adherence to global stages of development characteristic of most neo-Piagetian theories, Fischer and colleagues (1978, 1980a; 1981) advocate for ten levels through which skills develop, trifurcated into tiers: sensory motor/action (levels 1-3), representation (levels 4-6), and abstraction (levels 7-10). The levels in one tier repeat in the next tier, with the following distinctions: Level 1: a single set/control over one set of variation; Level 2: mapping; Level 3: a system; Level 4: systems of systems. The highest level of one tier (e.g., systems of systems in representation tier) is the lowest level in the next one (e.g., single set in the abstraction tier).

The theory organizes skills as hierarchies: lower-level skills are subsumed within higher levels. Individuals at higher levels for a particular skill can thus exhibit control over larger sources of variation, delineating increasing complexity of task performance. Five transformation rules[4]–describing how individuals coordinate skills to generate new skills at the same or higher levels–govern whether or not individuals progress through the levels within a skill domain (Fischer, 1980b, 1980a) . While the theory has many moving parts, the upshot is …"that development to higher levels is indicated by increasing complexity of tasks and performance in them" (Kallio, 2020, p. 18). In the proceeding section, we further highlight features of skill theory as they pertain to the focus of this study, connecting aspects of the theory to the development of intellectual abilities in emerging adulthood.

2.1 Is skill theory appropriate for studies of complex reasoning in higher education?

Skill theory provides a useful framework for the present study because: (a) the generality of the theory is inclusive of a broad range of complex reasoning skills; (b) equality is given to both the environment (e.g., college experiences) and the organism (e.g. student's individual characteristics and actions) in the formation of higher level skills; (c) the theory readily extends to emerging adulthood; and, finally, (d) the theory has been applied to develop and integrate some of the most popular models of cognitive and intellectual development in college students. We briefly expound on these points below, making clear the connection with contemporary higher education research.

Generality of cognition. For the first point, in skill theory, the "skill" remains abstract so as to generalize to a broad range of cognitive processes. Fischer makes this

[4] More information on the specifics of each transformation rule can be found in Fischer (1980a).

explicit: "…cognition includes anything that involves the person's controlling sources of variation, even when these sources have conventionally been called emotions, social skills, language, or whatever. All these various domains share the same processes of developing more and more effective cognitive control" (Fischer, 1980, p. 481).

Thus, skill theory is in no way limited to the domain of general intelligence and has been used across a wide-range of phenomena. In fact, it's been applied to areas as diverse as complex reasoning in young adulthood (King & Vanhecke, 2006; Kitchener & Fischer, 1990; Kitchener, Lynch, Fischer, & Wood, 1993; Schwartz & Fischer, 2003), identity development, (Kunnen & Bosma, 2003), social relationships (Lamborn, Fischer, & Pipp, 1994), moral reasoning (Glover & Steele, 1990; King, 2009), emotions and self-concept (Ayoub et al., 2006; Fischer, Shaver, & Carnochan, 1990; Harter, 1986), agency in pretend play (Chapman, 1987; Watson & Fischer, 1977) and even soccer expertise, measured as game-play evaluation/knowledge (Den Hartigh et al., 2014).

Equality of environment and organism. Because skills are defined as attributes of a person-in-a-context, Fischer and colleagues (1980a; 2014) avoid describing cognitive development as a linear one-track progression through universal stages. Instead, skill theory emphasizes the collaboration between organism and environment, highlighting how contextual support plays a major role in shaping the type and level of skills acquired, as well as how individuals may exhibit a swathe of growth trajectories bespeaking variability in cognitive profiles (Fischer & Granott, 1995; Fischer, Knight, & Van Parys, 1993). Environmental support includes opportunities for practice, prompting, observing skills modeled, and the provision of scaffolding (Bidell & Fischer, 1992; Fischer & Silvern, 1985).

A context can also elicit emotional states and affections that contribute to the development of skills (Fischer & Lamborn, 1989). Emotions affect the propensity to act in certain ways (e.g., action tendency); behavior in turn affects development via producing patterns of functioning (e.g., skills) in the emotion-producing context (Fischer et al., 1990; Fontaine & Scherer, 2013; Frijda, 1987). Moreover, environments can also interact with stable traits to spur cognitive processes (DeYoung, 2015; Hart, Keller, Edelstein, & Hofmann, 1998; Matthews, 1999). Both the environment and individual states and traits are therefore crucial in understanding skill acquisition (Fischer, Shaver, & Carnochan, 1989, 1990; Fischer & Yan, 2018). With this lens, the significance of everyday activities in college can be viewed as features of the context that are either supportive or non-supportive to student development (Bidell & Fischer, 1992; King & Kitchener, 1994; Schwartz & Fischer, 2003).

Connection to emerging adulthood. Skill theory intends to predict young adult cognition (Fischer, Yan, & Stewart, 2003). Levels nine and ten relate to the development of abstract systems and principles (systems of abstract systems), which is theoretically and empirically supported as occurring between the late teens and mid-twenties (Fischer, Hand, & Russell, 1984; Fisher & Bidell, 2006; King & Kitchener, 1994). Mascolo and Fischer (2015) further expound: "…18- to 20-year-olds are able to construct highly differentiated relations between multiple abstractions, each of which is grounded in the coordination of multiple systems of lower-level concrete representations of events. For example, a young adult can begin to represent the relation between two aspects of his career and two aspects of his personal life" (p. 125 - 126).

Arnett (2000) identified this same age group–ranging from 17 years of age to approximately the late twenties–as a unique period in life where individuals in Western industrialized societies undergo important changes. He argued that in *emerging adulthood* individuals are distinct in their demographic instability and self-concepts. Perhaps even more striking, Arnett (2000) also highlighted the intense identity exploration in love, work, and worldviews that accompanies this period of intermediate responsibility, fluid commitments, and exposure to diverse ideas and perspectives. Echoing research by Fischer and colleagues, emerging adulthood has since been viewed as "…as a period critical for the establishment of mature structures of thinking, yet also vulnerable to stabilizing distortive forms of thinking if important familial and cultural supports are not available" (Labouvie-Vief, 2006, p. 80).

Connection to models of intellectual development. Finally, in direct communication with the previous point, skill theory has been applied to develop and integrate some of the most popular models of cognitive and intellectual development among college students. For instance, King and Kitchner's (1994) model of reflective judgement–perhaps, the most researched and thus well-supported developmental model of emerging adult reasoning abilities–builds upon the theoretical foundations of skill theory. It informed the construction of the model itself, the accompanying assessment, and provided predictions for the sequence of age-related development in supportive and non-supportive contexts (Fischer & Pruyne, 2002; King & Kitchener, 1994; Kitchener & Fischer, 1990; Kitchener et al., 1993).

Moreover, and quite recently, skill theory has been shown to integrate a wide variety of neo-Piagetian emerging adult cognitive theories. Starting with the last two tiers (representation and abstraction) of skill theory (the first tier is omitted as these levels apply to infants), King and Kitchener (2015) map each level not only to the developmental stages in their own model of complex reasoning (reflective judgement), but to other models concerning epistemic cognition[5] or cognitive flexibility. For example, the positions in Perry's (1970) pioneering scheme of intellectual development, stages in Belenky et al's. (1986) *"Women's ways of Knowing"*, and 4-step growth pattern in Magolda Baxter's (Magolda Baxter, 1992) *Epistemological Reflection Model* were each shown to correspond with the general developmental trajectory of Fischer's (1980) theory, and could be mapped to various levels.

2.2 How does course-taking breadth relate to skill theory?

Course-taking breadth can be viewed as contextual support for the development of complex reasoning skills. First, if offers opportunities to practice the reasoning skills of interest in this study. With recent qualms over academic rigor in higher education (Arum & Roksa, 2011; Arum, Roksa, & Cook, 2016; Francis, 2018), there is a newfound emphasis on higher-order thinking skills. Students are often provided tasks that require processing and synthesizing information, identifying significant relationships, drawing conclusions, reflecting on diverse opinions, and developing and evaluating arguments (Barnett &

[5] Epistemic cognition is sometimes referred to as postformal: the stage beyond Piaget's formal reasoning. However, postformal reasoning has since become a more nuanced idea associated with similar theories of adult cognition (Sinnott, 1981, 1998, 2008). Nevertheless, these models deal largely with reasoning about ill-structured problems (problems that don't have clear solutions and require weighing uncertainties), and how beliefs about knowledge claims evolve over the course of attending college and progressing through the emerging adult years.

Francis, 2012; Brown & Bielinska-Kwapisz, 2015; Culver, Braxton, & Pascarella, 2019; Goodwin, Chittle, Dixon, & Andrews, 2018; Orona & Pritchard, 2021; Stanny, Gonzalez, & McGowan, 2015).

Second, it offers opportunities to practice these reasoning skills in new discipline areas. Norms from within a discipline are challenged by another; students must reconcile opposing vantage points and synthesize perspectives to generate judgements on a variety of dynamic issues (Kirk-Kuwaye & Sano-Franchini, 2015; Malamud, 2012). Course-taking breadth exposes students to more diverse perspectives (e.g., different theories; students from different majors and backgrounds; instructors with pedagogical styles and professional experiences) and should provide a context for dealing with ill-structured problems that arise either in graded assignments or class discussions.

Finally, even in large-scale lecture and online courses, instructors often model problem-solving in front of the class, use iClickers to check students understanding, encourage critical reflection, and provide opportunities for collaboration with peers for scaffolding (Orona, 2021a; Orona, Li, McPartlan, Bartek, & Xu, 2022; Radovic, Firssova, Hummel, & Vermeulen, 2021; Vu, 2017). These pedagogical techniques connect to supportive environmental features such as modelling appropriate behavior (e.g., reasoning skills), prompting the use of specific skills, and supplementing current abilities with targeted assistance.

With this supportive context, students are positioned to coordinate their skills in various ways, such as combining or substituting, to develop new skills at the same or higher level. King and Kitchener (1994, p. 35) discuss an example of skill acquisition: "…an individual may not be challenged to try to make sense of discrepancies between the competing views of different authorities because he or she has not previously been asked to compare or contrast such views. When the juxtaposition does occur, as may happen in an educational setting, it stimulates the person to work out a coordination of skills. As the skills are coordinated across levels, they are transformed into new skills. Without challenges that span different types of skills, a person's existing skill level will probably not change or develop into higher-order skills."

3. Current Study

The preceding section highlights the relevance dynamic skill theory has for studying cognitive development among emerging adults attending higher education institutions. In our view–given the rationale and literature outlined–the theory would predict a positive association between course-taking breadth and complex reasoning, though our credence is localized to modest parameter values tempered by previous research and the uncertainty of the quality of contextual support for inculcating reasoning abilities across task domains. Combining data collected from a university-wide measurement project and administrative records, we examine these relations. The current study thereby contributes to the literature in emerging adult development by examining how a longstanding tradition–course-taking breadth–and the interaction with key student dispositions (e.g., intellectual curiosity)–influence the development of complex reasoning skills spanning ill and well-structured problems in emerging adults.

4. Methods

4.1 Setting and Data Collection

This study takes place at a large public research university located in southern California. The university typically enrolls about 30,000 students each term. Approximately a month prior to the fall 2019 term, all entering freshman and returning juniors were asked to participate in the Measuring Undergraduate Success Trajectories project (Arum et al., 2021a, 2021b) via emails to their school email. Data collection began two weeks prior to the start of the fall 2019 quarter. The second wave of data collection occurred at the end of the spring 2021 term, whereby the same surveys and performance assessments were administered again among the same group of students.

In the first instance of data collection, the rationale for the study was relayed by highly esteemed professors and administrative leaders. Additionally, all surveys and performance assessments were administered with two graduate assistant proctors in each lecture room. These efforts were implemented to motivate students to complete the assessments to the best of their ability and elicit their optimal level of performance; studies suggest such techniques as effective in large-scale testing situations (Liu, Bridgeman, & Adler, 2012; Liu, Rios, & Borden, 2015; Rios, Guo, Mao, & Liu, 2017). In the second wave of data collection, all performance assessments were taken virtually due to the COVID-19 pandemic, and secure links were provided to participants during designated test-taking slots. Students were shown videos of the significance relaying of the study and online proctors monitored participants.

4.2 Participants

Two-hundred and sixty students completed all assessments at time 1 (fall 2019) and time 2 (spring 2021), had full data on the relevant demographic, academic, and survey variables, and were in the broad emerging adult age range (17-29). Table 3.1 displays the descriptive statistics of the sample. The mean age was 18.5. About 30% were underrepresented minority students (URM) and about the same proportion (29%) were male.

4.3 Measures

The independent variable of interest is out-of-major course-taking breadth (outer breadth). This measure was constructed by counting the number of courses student enrolled in that were not within a student's major requirements specific to that degree. For example, a biology student's outer breadth count would not include any biology courses, nor would it include required exchangeable possibilities, such as similar course(s) offered in the physical sciences. It would, however, include courses in the arts, humanities, social sciences, and other unrelated science, technology, engineering, and mathematics (STEM) fields. The average number of out-of-major course-taking was 11.18 (*SD* = 8.09), as depicted in table 3.1.

The reason a count is used as opposed to either a proportion or diversity index is because we are interested in examining if absolute exposure (not merely relative exposure) to the hypothesized supportive environment contributes to skill growth. For instance, suppose a student has taken 10 courses in the two-year time frame under study. If 5 of them are out-of-major, their outer breadth proportion is .5. Suppose another student has taken 24 courses, with 10 of them out-of-major. The latter student has a lower outer breadth proportion (10/24 = .41), yet double the amount of exposure to the hypothesized supportive condition. Such a measure examines if *balancing inner and outer breadth* leads to skill formation, which is not the focus of this study.

The primary dependent variable in this study is complex cognitive reasoning, defined as the ability to successfully navigate both well and ill-structured problems. To measure this broad construct–and avoid reliance on just one narrow type of assessment (Fischer, Stein, & Heikkinen, 2009)–we use four distinct cognitive scores. Together the assessment items span formal/logical operations (e.g., valid argumentation and deductive logic) and real-world evaluative and inference-making skills (e.g., biases in news sources); others emphasize capacities to understand and represent disparate views and require solutions to situations that have no right or wrong answer. We elaborate on each test and construct below.

Critical thinking. The HEIghten Critical Thinking (HCT) test is a performance assessment measuring critical thinking developed by the Educational Testing Service (ETS). Through an extensive review of the critical thinking literature, the test was designed to employ a range of item types and typically takes about 45 minutes to complete (Liu, Frankel, & Roohr, 2014; Liu, Mao, Frankel, & Xu, 2016). Two dimensions of critical thinking constitute the operational definition, confirmed in subsequent empirical work (Liu et al., 2016): analytic and synthetic reasoning. Analytic reasoning includes evaluating evidence and arguments. In this dimension, tasks are provided to test-takers that require judging the relevance of information/evidence as it relates to specific arguments, recognizing potential biases, and indicating the reliability of information sources. Synthetic reasoning includes understanding the implications and consequences of arguments and developing sound and valid arguments. In this dimension, tasks are provided to test-takers that require identifying appropriate conclusions supported by arguments, extrapolating implications to draw out a sequence, and identifying appropriate premises that lead to certain conclusions and vice-versa. The HCT has been validated in samples drawn from the United States (Liu et al., 2016), China (Liu et al., 2018), Russia (Shaw et al., 2020), and Ireland (O'Leary et al., 2020).

Civic online reasoning. Civic online reasoning (COR) is measured using the test developed by McGrew (2018; 2017), Wineburg (2016) and their colleagues. The test presents two prompts/items, each of which requires a written response from the student. The prompts juxtapose news articles and associated images taken from different sources (e.g., the Atlantic and a blog). Students are asked to rate their confidence in trusting each respective source by detailing their rationale for choosing one source as more trustworthy as the other. Responses are coded to examine (a) the correct choice and (b) thoughtfulness in why the choice was selected. The responses are scaled on a three-point scale, with 0 = Needing improvement, 1 = Intermediate, and 2 = Mastery.

Perspective-taking. The perspective-taking (PT) test is a novel assessment developed by ETS in collaboration with the authors. The PT test requires an essay-type response to a complex vignette. The vignette begins with a profile of two characters presented as students in higher education. Descriptions of their family backgrounds, personalities, general dispositions, aims and goals–all of which differ from one another–are provided. A short story is supplied whereby the two characters find themselves in a mildly confrontational situation. The test-taker is then asked to describe the dynamics of the confrontation and offer their suggestions on how it might be resolved. In so doing, students are scored on how well they represent the perspectives of both characters and whether their suggestions connect and are sensitive to the facts of the scenario and aims of the individual characters presented. Thus, they must articulate the views and rationale for the feelings and actions presented in the story without explicit direction to do so; additionally, test-takers are free to provide creative and innovative solutions. The responses are scaled on a five-point scale, with 1 = Insufficient, 0 = Limited, and 3 = Adequate, 4 = Strong, and 5 = Outstanding.

Complex cognitive reasoning. To test if these measures formed a broad skill domain (Fischer & Farrar, 1987), we specified a Bayesian structural equation model (BSEM) with each of the four assessment total scores serving as indicators for one general Complex Cognitive Reasoning (CCR) factor for each time point. For the model with no priors and no correlated uniqueness, the fit was acceptable: (Bayesian) Comparative Fit Index (BCFI) = 0.914; Tucker-Lewis Index (BTLI) = 0.872; Root Mean Square Error of Approximation (BRMSEA) = 0.077. Posterior means of the standardized factor loadings for the test scores were: analytic (T1: 0.71; T2: 0.73), synthetic (T1: 0.81; T2: 0.71), COR (T1: 0.47; T2: 0.44), and PT (T1: 0.33; T2: 0.41). Table 3.2 shows more fit indices alongside other model specifications using priors. The Bayesian greater lower bound (glb) reliability estimate was centered at 0.71 at time 1 and 0.67 at time 2, respectively. Based on this analysis, the four measures were subsequently standardized, summed, and standardized again to form an index of CCR for both assessment times (fall 2019 and spring 2021).

Control variables. The research question(s) in this study are causal–we do not wish to hide such language under the guise of description or prediction (Grosz, Rohrer, & Thoemmes, 2020). But since this study is nonexperimental, there are many threats to estimating a causal relationship. Specifically omitted variable bias, or selection into out-of-major course breadth. To think through the necessary statistical adjustments, we construct a directed acyclic graph (DAG) to depict the posited causal structure. DAGs are useful graphical tools that can assist researchers in constructing theory-driven models that not only include the relevant potential confounders, but also *exclude* mediators and colliders that may induce *included variable bias* (Rohrer, 2018; Wysocki, Lawson, & Rhemtulla, 2020; York, 2018).

In designing our DAG, we did not rely solely on skill theory: we utilized information from the broader empirical and theoretical literature regarding academic/intellectual experiences outcomes. For instance, we consulted the review by Mayhew, Rockenbach, Bowman, Seifert, and Wolniak (2016) to glean what college-experiences have been shown to affect cognitive abilities. Furthermore, in comparison to within-major courses, Lee, Keyserlingk, Arum, and Eccles (2021) showed that students choose to enroll in courses outside their major for a variety of reasons, including interest, a desire to broaden themselves, personal development, because they have the talent to succeed in the course, or because they perceive the course as easy. Additionally, Baker and Orona (2020) provide some evidence that demographics play a role in major awareness and consideration (which are related to major selection). And Ackerman's (1996) intelligence-as-process, personality, interests, and knowledge theory posits the relevance of investment traits and dispositions in pursuing specific intellectual endeavors.

Figure 3.1 shows our DAG. According to this presumed causal structure, adjusting for age, CCR1, SES, major, prior academic achievement (PA1), extra-curricular activities (ECA1)–like honor programs, major, residence (RES1) and individual dispositions and attributes (INT1), such as intellectual curiosity and the tendency to exert effort, closes all backdoor paths that would transmit a spurious correlation between outer breadth (OB) and CCR2. Of course, our causal structure is more than likely incorrect and there are potentially other unmeasured variables obfuscating the relationship between outer breadth and CCR2. We therefore focus on the full model, yet test multiple models (including the model represented in the DAG) to examine the stability of parameter values.

4.4 Prior elicitation

Per contemporary recommendations (e.g., Zondervan-Zwijnenburg, Peeters, Depaoli, & Van de Schoot, 2017), our priors are informed by the relevant empirical research. For instance, Bangert-Drown's (1990) meta-analysis showed that educational interventions have, on average, an effect on critical thinking of about .37 standard deviation units (effect size). Arum and Roksa (2011) found a critical thinking change of .18 in standard deviation units over three college semesters. In their meta-analysis, Huber and Kuncel (2016) found an average effect size of .46 over four years of college. Wang, Pascarella, Nelson Laird, and Ribera (2015) found that greater exposure to reflective thinking was associated with critical thinking at .13; though they also found negative associations for higher-order learning and integrative learning (-0.01 and -0.042), respectively. Guerra-Carrillo, Katovich, and Bunge (2017) found that two years of post-secondary education is associated with a .14 gain in cognitive performance. Blaich, Wise, and Crawfordsville (2011) found a .44 average change in critical thinking over the course of four years of college. Finally, King and Kitchener (1994) found a approximate standardized difference of .09 between juniors and freshman in reflective thinking scores. More recently, and using at least one common assessment constituting the complex cognitive reasoning construct, Liu, Roohr, Seybert, and Fishtein (2021) found a cross-sectional standardized effect size difference of .24 between seniors and freshman in their overall HEighten critical thinking scores (HCT) across all majors. They also found that those exposed to a higher frequency of applying critical thinking skills in courses exhibited greater HCT scores (effect size = .25) than those who were only sometimes exposed.

In generating our prior based on this research, there are several things that must be considered. First, there are time differences. While we can try to account for this with simple calculations (e.g., a 4-year change can be divided by 2 to reflect a 2-year change), this assumes (perhaps erroneously) that change is equal across time intervals. Second, effect sizes need to be converted to approximate standardized regression coefficients. We do this by following the guidelines suggested by Mayhew et al. (2016, p. 20) in that a standardized effect size multiplied by .4 provides an approximation to an equivalently sized regression weight. After considering these two points, and naively weighing all studies equally, a simple mean of previous associations is .09.

However, additional concerns, liked the fact that no study looked at the exact same reasoning construct, nor examined the same predictor, nor specified the same statistical model as the present study leaves this estimate wanting. Additionally, the theoretical support for conceiving out-of-major breadth as a supportive context for reasoning abilities leads us to construct a prior that is centered on a bit larger value: .13. But because of the uncertainty regarding this estimate, we specify a standard deviation of .2. More information on how the prior is integrated into the statistical model, how we check it to examine its plausibility, and its effect on the results can be found in the data analysis, prior predictive checking, and sensitivity analysis sections below.

4.5 Data analysis

We employ Bayesian regression to specify the relationship between course breadth and CCR. The full model includes several covariates and is described below:

$$Y_t \sim Normal(\mu_t, \sigma)$$

$$\mu_i = \alpha + \beta_1 CourseTakingBreadth + \beta_2 CCR1 + \sum_{j=1}^{n=14} \beta_j X_{ji}$$

$$\alpha \sim Normal(0, 2.5)$$

$$\beta_1 \sim Normal(.13, .2)$$

$$\beta_2 \sim Normal(.7, .5)$$

$$\beta_j \sim Normal(0, 1)$$

$$\sigma \sim Exponential(1)$$

where Y_{it} represents the outcome (CCR2) for every student i; α is the constant; β_1 is the parameter representing the association between course-taking breadth and CCR2; β_2 represents the initial measure of CCR at time point 1; and β_j indicates a vector of student-level control variables, including: student major, URM status, class standing (freshman or junior), age, gender, high school GPA, SAT cumulative score, socioeconomic status, full-time status, curiosity–a composite generated from openness to experience, epistemic curiosity, and need for cognition–(Powell, Nettelbeck, & Burns, 2016; Von Stumm, Hell, & Chamorro-Premuzic, 2011), effort–a composite generated from conscientiousness, academic self-efficacy, and selective primary control–(Fagioli, Baker, & Orona, 2020; Heckhausen, Schulz, & Wrosch, 1999; Orona, 2021b), political and social involvement (Blaich et al., 2011), and one item relating to an individuals' utility over monetary returns in future employment and a desire for general education, respectively.

5. Results

We organize our results section using the WAMBS-v2 (Depaoli & Van de Schoot, 2017; Van de Schoot et al., 2021; Van de Schoot, Veen, Smeets, Winter, & Depaoli, 2020). The WAMBS (When to worry and how to Avoid the Misuse of Bayesian Statistics) checklist specifies 10 points. We review these now.

5.1 Point 1: Prior Predictive Checking

Step 1 of the WAMBS encourages prior predictive checking, which is a way of evaluating our credence about likely parameter values by examining the kinds of predictions they make. In lay terms, this could be viewed as a self-check, asking the question: "Is this really your best guess?" "Do you *really* consider these predictions plausible, and to the extent shown?" In figure 3.2, we show simulated distributions of parameter values based upon the priors specified for every parameter in the model. To repeat, these distributions are not based on the study data; they are simulated from our priors. As can be viewed, all distributions appear reasonable and consistent with our beliefs: most parameters are centered at or close to 0, with large values having very low probability mass. For our primary variable of interest–out-of-major-breadth–we see a distribution centered at .13; much larger values in either direction (-0.5 and 0.5) are at the very ends of the distribution. Therefore, we do not view our prior specifications as making wild predictions inconsistent with our level of information.

5.2 Point(s) 2-5: Convergence Diagnostics

Convergence diagnostics are used to show whether or not the Markov Chain Monte Carlo (MCMC) iterations show evidence of reliable sampling from the posterior distribution(s). There are several ways of assessing convergence. In figure 3.3, we display the traceplots for every parameter in the model. We see that each parameter displays well-mixed chains and stationarity (iterations localized to a general area). Furthermore, the Gelman-Rubin convergence diagnostic (Gelman & Rubin, 1992), $\hat{R}$, for every parameter is 1.

Also the number of effective sample sizes sampled for every parameter is not demonstrably lower than the number of iterations. Finally, the shape of each parameter is normally distributed (as specified our model and presented in figure 3.4), suggesting a sufficient number of samples. Together, these criteria suggest convergence.

5.3 Point 6: Posterior Predictive Checking

Posterior predictive checking, unlike prior predictive checking, simulates data based on the specified model and compares the simulated data (*yrep*) to the actual data (*y*). Figure 3.5 shows the posterior distribution of the model compared against the many replications from the model. We see, generally, an acceptable looking fit of the model to the data, as most replications tightly surround the data distribution. There is a replication that appears noticeably more peaked than the data, which should not be too alarming.

5.4 Point(s) 7-9: Sensitivity Analysis

One major critique of Bayesian statistics is that it is subjective, incorporating and encoding researchers' own beliefs in a statistical model. The justification for priors and a defense of subjective probability more generally is beyond the purview of this study. But one recurring concern is if priors determine results. While we believe that this somewhat misses the ethos of Bayesian epistemology, a sensitivity analysis, however, can be very useful. A sensitivity analysis is the specification of different priors to examine both the stability of the estimates and the influence of the original priors on the model results. In our view, an additional benefit of sensitivity analysis not mentioned in the literature is that it can be used to encode *other beliefs*–that is, beliefs held by individuals other than the researcher(s) conducting the study. In this way, an array of beliefs can be represented and subsequently updated simultaneously.

Table 3.4 displays the results of the sensitivity analysis. We showcase 8 different priors (including the original priors) for the out-of-major breadth (OuterBreadth) parameter. From the top, the first prior results in a posterior mean and standard deviation nearly identical to the prior–this prior encodes a dogmatic belief: there was virtually no updating. As we move down the rows in table 3.4, the prior become more relaxed. It can be viewed that the most agnostic or weakly informative models result in larger posterior means.

5.5 Points 10: Bayesian Interpretations

Table 3.3 shows the results of the Bayesian regression model. The results show a posterior mean centered on a moderate-to-strong correlation value of .17, with 89% of the high-density interval spanning parameter values starting at .05 to .29. These results suggest that our credence is centered at a moderate-to-strong value; moreover, the evidence suggests quite strongly that the probability of a positive association between course-taking breadth and CCR is highly likely, with approximately 98% of the posterior probability mass covering positive parameter values.

In figure 3.6, we present the prior (pink line), posterior (blue line), and likelihood (black line). We see that our beginning credence was–although quite informative–relatively wide–and centered over .13, as detailed in the analytic plan. Once the model was specified, the information in the data (e.g., the likelihood) suggests a distribution centered at a about

.195. Per Baye's Theorem, the posterior is a function of of the prior $Pr(\theta)$ and the data $Pr(y|\theta)$ and can be written as:

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)}$$

thus, our posterior in figure 3.6 shows how it is influenced by (in the middle of) both the prior and likelihood, centered at .17. Since there is more uncertainty in our prior relative to the strength of the data, our posterior is more heavily influenced by the data than our prior.

Finally, figure 3.7 depicts the partial regression plot between outer breadth and CCR2 (complex cognitive reasoning at time 2), controlling for all other variables in the model. Visually, there appears to be some evidence of a linear trend, though it appears moderate at best, as the point estimate suggests.

5.6 Individual Test Results

The same procedures were followed for each individual cognitive test, and the same prior over the outer breadth parameter used in the previous model was also used here. Table 3.5 shows the posterior summary for each of the individual performance assessments using informative priors. A sensitivity analysis using default priors is also shown. Figure 3.8 shows the marginal posterior distributions for the relation between outer breadth and each cognitive test. It appears that most of the probability mass is over positive values for all four assessments, with the relations with analytic and synthetic dimensions having the largest posterior means. Outer breadth was most weakly related to

PT; this parameter also had widest distribution of the four (posterior *mean* = .08, posterior *SD* = .084).

5.7 Environmental and Organismic Interactions

Environmental forces do not spur development in isolation; stimuli interact with attributes of individual organisms. In this study, we focus on two attributes previously identified as highly salient in academic learning contexts: intellectual curiosity and the control of and tendency to exert effort (Von Stumm, 2013; Von Stumm et al., 2011). As mentioned in the measures section, both dispositions were measured with several self-report scales to generate composites. IC interacted with outer breadth was located on a sizeable correlation value, while effort did not (0.13 and 0.01, respectively.

As continuous-by-continuous interactions can be difficult to interpret, we provide figure 3.9 which depicts IC and effort into as groupings. We see those students +1SD above the mean level of IC and low on outer breadth are low on CCR2; however, students +1SD above the mean who are also high on outer breadth are higher on CCR2 than those at or -1SD below the mean. With effort, there doesn't appear to be strong evidence of interaction, either graphically or based upon the posterior distribution of the interaction term.

5.8 Robustness Checks

Just more courses? How do we know that out-of-major breadth indeed supports reasoning skill formation or if the association is merely driven by taking more courses, regardless of whether they are inside or outside ones' major? To test if outer breadth has a unique supporting role in skill development, we examine the direct influence of both total courses taken and inner breadth [6] (breadth within a major). Table 3.6 displays the results of the robustness check. The posterior mean for both total courses taken and inner breadth (within major courses taken) are much smaller than that for outer breadth (table 3.2); in fact, the inner breadth posterior mean is negative (-0.16). Moreover, both of the 89% credible intervals include negative values and are wider (bespeaking less confidence) than that of outer breadth. Together, these two additional analyses relay that neither total courses taken, nor more courses taken within one's major has the same positive association with complex cognitive reasoning as does outer breadth. The full posterior distributions for these relationships can be found in figure 3.10.

Measurement and latent variable models. As another robustness check, we specify a Bayesian structural equation model (BSEM) to account for measurement error in in the CCR variables. We use each test score as an indicator. The posterior mean and standard deviation was .18 and .08, respectively. The distributions from the composite and measurement models were therefore similar.

Multiple models. As depicted in figure 1, we presented a hypothesized causal structure. We compare this structure against a variety of models, including the full model. Table 3.7 shows the posterior summary and the out-of-sample predictive criteria for six different models. The top row—the DAG model—only includes predictors that close the backdoor paths presented in figure 1. These include: CCR1, age, attitude/disposition

[6] Inner breadth is used instead of the term *depth* because depth concerns how deep a student goes into a particular topic/subject area. Here, the inner breadth measure constitutes how many courses students take in their major, regardless if any one topic is explored in detail.

variables, major, extra-curricular honor's program, and prior academic achievement. The second row strips all predictors and only examines outer breadth with CCR2. From there, academic, demographic, and attitude/dispositional variables are added, ultimately leading to the full model. It can be viewed that as controls are added, the posterior mean of β_1 increases. Additionally, the Watanabe-Akaike information criterion (WAIC) was lowest for the model with only the pretest and academic (major, high school GPA, and SAT scores) variables, bespeaking the most parsimonious model of this limited set of models. Interestingly, this model generated a β_1 posterior mean of .11; the DAG model displayed a posterior mean of .15, and the full model, already reviewed in the main analysis, had a posterior mean of .17. While the variation is clear, the estimates are not excessively unwieldy; there is much overlap in the 89% credible intervals.

6. Discussion

Kuhn (1999, p. 16) noted, "…the burgeoning critical thinking movement in education has proceeded with little apparent contribution from contemporary cognitive development research." Recognizing this shortcoming some 20 years later, the current study applied dynamic skill theory–a robust and enduring cognitive theory–to examine the relationship between course-taking breadth and complex reasoning, connecting cognitive development research to higher education assessment. The results have several implications for theory, research, and practice.

6.1 Theoretical Implications

Fischer and Kenny (1986) suggested that at minimum skill building requires: (a) an environment that supports behaviors related to the skill in question and (b) an individual who takes advantage of the context and practices the skill. The results suggest that, in accordance with our predictions grounded in Fischer's skill theory, taking courses out of one's major provides a supportive context for the exercise and eventual development of complex reasoning skills.

Interestingly, while there was noticeable probability mass over sizable parameter values for the overall measure, it appears that this context is more conducive to building critical thinking skills (e.g., analytic and synthetic reasoning), than civic online reasoning and perspective-taking. Of the four measures constituting complex cognitive reasoning, outer breadth had the weakest correlation with perspective-taking. One very plausible explanation for this is that, given this study took place during the COVID-19 pandemic and most courses were taken online, the environment (online courses) provided weak support for building this skill, while the others were less affected. It makes sense that interacting with people directly as opposed to only via discussion boards boosts perspective-taking skills to a greater degree. Future research is needed to test these suppositions.

Moreover, relevant individual attributes, such as intellectual curiosity, appear to moderate the influence of the context such that outer breadth furnishes more reasoning skills for those who are sufficiently curious. The moderating role of intellectual curiosity not only reinforces Fischer's notion of the significance of environmental and organismic interactions, but also connects with recent work aiming to enhance character attributes conducive to clear thinking (Orona, 2021c, e.g., intellectual virtues/cognitive character). If curricular experiences–such as course-taking breadth–interact with curiosity to produce cognitive gains, then researchers can begin to investigate dynamic pedagogical innovations targeting dispositions and traits. Such work is already underway. Orona and Pritchard (2021), for instance, introduce a novel online module aimed at increasing students' intellectual curiosity; the pilot study suggested preliminary effectiveness. Additionally, Orona (2021b) explored the possible mechanisms by which these traits develop.

The positive association between course breadth and complex reasoning depended much less on the effort construct. Associated constructs (e.g., conscientiousness) have shown to moderate the effect of cognitive abilities, like intelligence (Beaujean et al., 2011) in predicting academic achievement. While our finding is somewhat counter to expectations, it may be that raw determination and tenacity–while important–are not the attributes necessary to make reasoning gains in the *context of course breadth*. Future work should continue to investigate more intellectually oriented tenacity constructs and their association with educational environments and cognitive outcomes.

Another interesting facet of these results under the lens of skill theory is attributing growth to either the development of a new level or generalization. Generalization is when a skill(s) in one domain can be transferred to another because of the similarity between the two. Fischer and Farrar (1987) distinguish true from apparent generalization by noting that when individuals reach a new upper limit (in the abstraction tier, for instance), they will show synchronous spurts across a broad range of domains that are seemingly unrelated (e.g., relations between mathematical entities and personality conflict both require coordinating abstractions), which obfuscates whether generalization took place (Fischer & Farrar, 1987; Fischer, Kenny, & Pipp, 1990).

For the present situation, the question is: Did outer breadth induce analytic reasoning skills *only*, and students subsequently generalized those skills to synthetic, civic online, and perspective-taking tasks, or did they instead obtain a new optimal-level that allowed for growth across a broad range of tasks? Furthermore, it's possible that the environmental stimulus (outer breadth) influenced each task independently, and neither cognitive process occurred (Fischer & Farrar, 1987). While our data do not allow for us to distinguish true from apparent generalization, the fact that the *assessment conditions*–while aiming to glean maximum performance–did not provide aid and support, means that *functional* as opposed to *optimal-level* was assessed, and therefore the gains observed are more than likely not attributed to students moving to new levels of abstraction (Fischer, 1983, 1987; Fischer & Pipp, 1984). In line with recent critical thinking frameworks (Dwyer, Hogan, & Stewart, 2014) suggesting the plausibility of logical/argumentation skills co-occurring with the development of perspective-taking and other ill-structured problem-solving abilities (e.g., reflective judgement), adult cognitive developmental researchers can begin to explore this largely uncharted terrain.

6.2 Policy Implications

This study has implications for *policy research*, but not *policy-making*. The distinction is profound. Our results offer probabilistic statements about the relationship between course-taking breadth and complex reasoning but say nothing with respect to whether or not an *investment in course-taking breadth is worth the cost*. To inform decisions, we would need to overlay our posterior distribution with utilities (Atkinson & Dorfman, 2005; Lindley, 2000; Savage, 1972). In other words, we would need to combine the probabilities from this study with the cost-benefit of educational breadth to determine whether it's an avenue that warrants continued or expanded funding. It's completely feasible that the cognitive gains found in this study do not have enough practical value–enough real-world bang for the buck. Thus, policy research is needed to investigate this important question. Apropos to cost-benefit, it's unclear whether the complex reasoning gains found in this (and other) studies are worth the possibility of *forgone domain knowledge*.

6.3 Limitations and Future Research

There are several limitations to this research. Regarding external validity, the study university is in a heterogeneous region and enrolls a diverse student body, but future research can attempt a multi-site study spanning several large public universities to enhance generalizability. Additionally, while a strength of this study is the deployment of several performance assessments spanning related yet distinct individual constructs, eschewing reliance on any one assessment tool or item format, our study does not attempt and is not equipped to make appellations regarding stages of development. Relatedly, the primary construct of interest–complex cognitive reasoning–does not have a corresponding model describing the specific and sequential steps of development. While this weakness does not compromise the study results, it does somewhat delimit our theoretical understanding of the significance of the strength of the relationships observed.

And finally, this study is nonexperimental: students were not randomized to outer breadth exposure. We do, however, emphasize several salient study features highlighted by McShane, Gal, Gelman, Robert, and Tackett (2019) that increase our confidence in the validity of the results. First, with the advent of the replication crisis and the growing concern over the use of *p*-values (McShane et al., 2019; Wasserstein & Lazar, 2016), we abandon their use in this study, instead focusing on credible intervals and entire posterior distributions. Second, we use prior evidence to regularize the correlation of interest; this regularization served to *downwardly* adjust the posterior distribution for outer breadth. Third, the plausibility of the mechanism linking outer breadth and cognitive gain is high. The notion that a broad, general education (a): exposes individuals to new ideas and concepts, (b) requires them to analyze, synthesize, and integrate these concepts to (c) develop new abstractions that leads to (d) clearer thinking is not only motivated by the cognitive theory presently applied, but has been posited as the benefit of liberal education for centuries (Newman, 1852/1982). Fourth, the study design utilizes a longitudinal data collection procedure, enabling statistical adjustment of the time 1 pretest, alongside a host of theoretically relevant demographic, academic, and attitudinal measures (Wysocki et al., 2020).

The fifth relevant study factor regards data quality. Our outcome was measured via a mix of cognitive performance assessments and showed a suitable fit as a latent variable. Failing to specify a measurement model can result in bias estimates (Bollen & Lennox, 1991; Westfall & Yarkoni, 2016), and specifying a measurement model when one shouldn't can result in even greater bias (Rhemtulla, Bork, & Borsboom, 2020). Given the absence of a strong theoretical basis for the construct under study, the fact that both the latent variable and composite multiple regression analysis largely converged is regarded as evidence of stability across measurement specifications. Additionally, our primary independent variable was constructed via institutional records. With concerns over the validity of self-report instruments (Fagioli et al., 2020; Maul, 2017), we view these aspects of the current study as advantageous.

6.4 Conclusion

In this study, we began by introducing and arguing for the relevance of Fischer's dynamic skill theory in studying cognitive development among emerging adults. Connecting the longstanding liberal arts tradition of broad learning to reinforcing environmental conditions, we then tested how course-taking breadth buttresses the development of complex cognitive reasoning measured via a mix of well and ill-structured tasks. The implications of these results reaffirm skill theory's applicability in emerging adulthood and connects to critical thinking assessment studies in higher education. Given that many cognitive skills are highly relevant for competing in the modern workforce, sifting through misinformation, and pursuing an eudemonic life, the onus is on researchers to continuously vet supportive educational environments that enhance cognitive skills.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227–257.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469.

Arum, R., Eccles, J. S., Heckhausen, J., Orona, G. A., Keyserlingk, L. von, Wegemer, C. M., … Yamaguchi-Pedroza, K. (2021a). A framework for measuring undergraduate learning and growth. *Change: The Magazine of Higher Learning*, *53*(6), 51–59.

Arum, R., Eccles, J. S., Heckhausen, J., Orona, G. A., Keyserlingk, L. von, Wegemer, C. M., … Yamaguchi-Pedroza, K. (2021b). Ensuring a more equitable future: Assessing student learning and growth in higher education. *Postsecondary Value Commission*.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. University of Chicago Press.

Arum, R., Roksa, J., & Cook, A. (2016). *Improving quality in american higher education: Learning outcomes and assessments for the 21st century*. John Wiley & Sons.

Arum, R., & Stevens, M. L. (2020). Building tomorrow's workforce today: Twin proposals for the future of learning, opportunity, and work. *The Hamilton Project, Brookings Institution, Washington, DC*.

Atkinson, S. E., & Dorfman, J. H. (2005). Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: Crediting electric utilities for reducing air pollution. *Journal of Econometrics*, *126*(2), 445–468.

Ayoub, C. C., O'connor, E., Rappolt-Schlichtmann, G., Fischer, K. W., Rogosch, F. A., Toth, S. L., & Cicchetti, D. (2006). Cognitive and emotional differences in young maltreated children: A translational application of dynamic skill theory. *Development and Psychopathology*, *18*(3), 679–706.

Baker, R., & Orona, G. A. (2020). Gender and racial differences in awareness and consideration of curricular programs: Exploring a multistage model of major choice. *AERA Open*, *6*(3), 2332858420937023.

Bangert-Drowns, R. L., & Bankert, E. (1990). Meta-analysis of effects of explicit instruction for critical thinking.

Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, *32*(2), 201–211.

Beaujean, A. A., Firmin, M. W., Attai, S., Johnson, C. B., Firmin, R. L., & Mena, K. E. (2011). Using personality and cognitive ability to predict academic achievement in a young adult sample. *Personality and Individual Differences*, *51*(6), 709–714.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind* (Vol. 15). Basic books New York.

Bidell, T. R., & Fischer, K. W. (1992). Cognitive development in educational contexts: Implications of skill theory. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-piagetian theories of cognitive development: Implications and applications for education* (pp. 11–30). Routledge & Kegan Paul London.

Blaich, C., Wise, K., & Crawfordsville, I. (2011). The wabash national study: The impact of teaching practices and institutional conditions on student growth. In *Annual meeting of the american educational research association, new orleans, LA*. Citeseer.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305.

Brown, F. W., & Bielinska-Kwapisz, A. (2015). Understanding the nature and determinants of critical thinking among senior business undergraduate students. *Journal of Education for Business*, *90*(7), 359–368.

Chapman, M. (1987). A longitudinal study of cognitive representation in symbolic play, self-recognition, and object permanence during the second year. *International Journal of Behavioral Development*, *10*(2), 151–170.

Culver, K., Braxton, J., & Pascarella, E. (2019). Does teaching rigorously really enhance undergraduates' intellectual development? The relationship of academic rigor with critical thinking skills and lifelong learning motivations. *Higher Education*, *78*(4), 611–627.

Den Hartigh, R. J., Van Der Steen, S., De Meij, M., Van Yperen, N. W., Gernigon, C., & Van Geert, P. L. (2014). Characterising expert representations during real-time action: A skill theory application to soccer. *Journal of Cognitive Psychology*, *26*(7), 754–767.

Depaoli, S., & Van de Schoot, R. (2017). Improving transparency and replication in bayesian statistics: The WAMBS-checklist. *Psychological Methods*, *22*(2), 240.

DeYoung, C. G. (2015). Openness/intellect: A dimension of personality reflecting cognitive exploration. In *The APA handbook of personality and social psychology: Personality processes and individual differences* (pp. 369–399). American Psychological Association.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52.

Fagioli, L. P., Baker, R., & Orona, G. A. (2020). The role of non-cognitive variables in identifying community college students in need of targeted supports. *Research in Higher Education*, 1–39.

Fischer, K. W. (1978). Structural explanation of developmental change. *Behavioral and Brain Sciences*, *1*(2), 186–187.

Fischer, K. W. (1980a). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*(6), 477.

Fischer, K. W. (1980b). Learning as the development of organized behavior. *Journal of Structural Learning*.

Fischer, K. W. (1983). Developmental levels as periods of discontinuity. *New Directions for Child and Adolescent Development*, *1983*(21), 5–20.

Fischer, K. W. (1987). Relations between brain and cognitive development. *Child Development*, 623–632.

Fischer, K. W., & Bullock, D. (1981). Patterns of data: Sequence, synchrony, and constraint in cognitive development. *New Directions for Child and Adolescent Development*, *1981*(12), 1–20.

Fischer, K. W., Bullock, D. H., Rotenberg, E. J., & Raya, P. (2014). The dynamics of competence: How context contributes directly to skill. In *Development in context* (pp. 109–134). Psychology Press.

Fischer, K. W., & Farrar, M. J. (1987). Generalizations about generalization: How a theory of skill development explains both generality and specificity. *International Journal of Psychology*, *22*(5-6), 643–677.

Fischer, K. W., & Granott, N. (1995). Beyond one-dimensional change: Parallel, concurrent, socially distributed processes in learning and development. *Human Development*, *38*(6), 302–314.

Fischer, K. W., Hand, H. H., & Russell, S. (1984). The development of abstractions in adolescence and adulthood. In M. S. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations: Late adolescent and adult cognitive development* (pp. 43–73). New York: Prager.

Fischer, K. W., & Kenny, S. (1986). Environmental conditions for discontinuities in the development of abstractions. In *Adult cognitive development: Methods and models* (pp. 57–75). Praeger New York.

Fischer, K. W., Kenny, S. L., & Pipp, S. L. (1990). How cognitive processes and environmental conditions organize discontinuities in the development of abstractions. In C. Alexander & E. J. Langer (Eds.), *Higher stages of human development: Perspectives on adult growth*. Oxford University Press.

Fischer, K. W., Knight, C. C., & Van Parys, M. (1993). Analyzing diversity in developmental pathways: Methods and concepts. *Constructivist Approaches to Development: Contributions to Human Development*, *23*, 33–56.

Fischer, K. W., & Lamborn, S. D. (1989). Mechanisms of variation in developmental levels: Cognitive and emotional transitions during adolescence. In A. de Ribaupierre (Ed.), *Transition mechanisms in child development: The longitudinal perspective* (pp. 33–67). Cambridge University Press.

Fischer, K. W., & Pipp, S. L. (1984). Processes of cognitive development: Optimal level and skill acquisition. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 45–80).

Fischer, K. W., & Pruyne, E. (2002). Reflective thinking in adulthood. In *Handbook of adult development* (pp. 169–198). Springer.

Fischer, K. W., Shaver, P. R., & Carnochan, P. (1989). A skill approach to emotional development: From basic-to subordinate-category emotions. In *Child development today and tomorrow* (pp. 107–136). San Francisco: Jossey-Bass.

Fischer, K. W., Shaver, P. R., & Carnochan, P. (1990). How emotions develop and how they organise development. *Cognition and Emotion*, *4*(2), 81–127.

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology*, *36*(1), 613–648.

Fischer, K. W., Stein, Z., & Heikkinen, K. (2009). Narrow assessments misrepresent development and misguide policy: Comment on steinberg, cauffman, woolard, graham, and banich (2009). *American Psychologist*, *64*(7), 595–600.

Fischer, K. W., & Yan, Z. (2018). The development of dynamic skill theory. In *Conceptions of development* (pp. 279–312). Psychology Press.

Fischer, K. W., Yan, Z., & Stewart, J. (2003). Adult cognitive development: Dynamics in the developmental web. In J. Valsiner & K. Connolly (Eds.), *Handbook of developmental psychology* (pp. 491–516). Citeseer.

Fisher, K. W., & Bidell, T. (2006). Dynamic development of action, thought and emotion. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development: Handbook of child psychology* (pp. 313–399). New York: Wiley.

Fontaine, J., & Scherer, K. (2013). Emotion is for doing: The action tendency component. In *Components of emotional meaning: A sourcebook* (pp. 170–185). Oxford University Press.

Francis, C. (2018). Academic rigor in the college classroom: Two federal commissions strive to define rigor in the past 70 years. *New Directions for Higher Education*, *2018*(181), 25–34.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*(2), 115–143.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Glover, R. J., & Steele, C. (1990). Applying neopiagetian theory to the moral reasoning process. *Psychological Reports*, *66*(3), 1259–1272.

Goodwin, A., Chittle, L., Dixon, J. C., & Andrews, D. M. (2018). Taking stock and effecting change: Curriculum evaluation through a review of course syllabi. *Assessment & Evaluation in Higher Education*, *43*(6), 855–866.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255.

Guerra-Carrillo, B., Katovich, K., & Bunge, S. A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. *PloS One*, *12*(8), e0182276.

Hart, D., Keller, M., Edelstein, W., & Hofmann, V. (1998). Childhood personality influences on social–cognitive development: A longitudinal study. *Journal of Personality and Social Psychology*, *74*(5), 1278.

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition*, *4*(2), 119–151.

Heckhausen, J., Schulz, R., & Wrosch, C. (1999). *Optimization in primary and secondary control (OPS-scales)*. Technical report. Max Planck Institute for Human Development; Education ….

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, *86*(2), 431–468.

Kallio, E. K. (2020). From multiperspective to contextual integrative thinking in adulthood: Considerations on theorisation of adult thinking and its place as a component of wisdom 1. In *Development of adult thinking* (pp. 9–32). Routledge.

King, P. M. (2009). Principles of development and developmental change underlying theories of cognitive and moral development. *Journal of College Student Development*, *50*(6), 597–620.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco: Jossey-Bass.

King, P. M., & Kitchener, K. S. (2015). Cognitive development in the emerging adult: The emergence of complex cognitive skills. *The Oxford Handbook of Emerging Adulthood*, 105–125.

King, P. M., & Vanhecke, J. R. (2006). Making connections: Using skill theory to recognize how students build and rebuild understanding. *About Campus*, *11*(1), 10–16.

Kirk-Kuwaye, M., & Sano-Franchini, D. (2015). "Why do i have to take this course?" How academic advisers can help students find personal meaning and purpose in general education. *The Journal of General Education*, *64*(2), 99–105.

Kitchener, K. S., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. *Contributions to Human Development*.

Kitchener, K. S., Lynch, C. L., Fischer, K. W., & Wood, P. K. (1993). Developmental range of reflective judgment: The effect of contextual support and practice on developmental stage. *Developmental Psychology*, *29*(5), 893.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, *28*(2), 16–46.

Kunnen, E. S., & Bosma, H. A. (2003). Fischer's skill theory applied to identity development: A response to kroger. *Identity: An International Journal of Theory and Research*, *3*(3), 247–270.

Labouvie-Vief, G. (2006). Emerging structures of adult thought. In *Emerging adults in america: Coming of age in the 21st century.* (pp. 59–84). American Psychological Association.

Lamborn, S. D., Fischer, K. W., & Pipp, S. (1994). Constructive criticism and social lies: A developmental sequence for understanding honesty and kindness in social interactions. *Developmental Psychology*, *30*(4), 495.

Lee, H. R., Keyserlingk, L. von, Arum, R., & Eccles, J. S. (2021). Why do they enroll in this course? Undergraduates' course choice from a motivational perspective. *Frontiers in Education*, *6*, 1–9.

Lindley, D. V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(3), 293–337.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, *41*(9), 352–362.

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, *2014*(1), 1–23.

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education*, *41*(5), 677–694.

Liu, O. L., Rios, J. A., & Borden, V. (2015). The effects of motivational instruction on college students' performance on low-stakes assessment. *Educational Assessment*, *20*(2), 79–94.

Liu, O. L., Roohr, K. C., Seybert, J. M., & Fishtein, D. (2021). Are fourth-year college students better critical thinkers than their first-year peers? Not so much, and college major and ethnicity matter. *Educational Measurement: Issues and Practice*.

Liu, O. L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., … others. (2018). Assessing college critical thinking: Preliminary results from the chinese HEIghten critical thinking assessment. *Higher Education Research & Development*, *37*(5), 999–1014.

Magolda Baxter, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development.* Jossey-Bass.

Malamud, O. (2012). The effect of curriculum breadth and general skills on unemployment. In *Society of labor economists conference* (pp. 4–5).

Mascolo, M. F. (2020). Dynamic skill theory: An integrative model of psychological development. In *Handbook of integrative developmental science* (pp. 91–135). Routledge.

Mascolo, M. F., & Fischer, K. W. (2015). Dynamic development of thinking, feeling, and acting. In W. F. Overton, P. C. M. Molenaar, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science* (pp. 115–161). Wiley Online Library.

Matthews, G. (1999). Personality and skill: A cognitive-adaptive framework. In *Learning and individual differences: Process, trait, and content determinants.* (pp. 251–273). American Psychological Association.

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 51–69.

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A., & Wolniak, G. C. (2016). *How college affects students: 21st century evidence that higher education works* (Vol. 1). John Wiley & Sons.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, *46*(2), 165–193.

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Teaching students to engage in civic online reasoning. *American Educator*, *41*(3), 4.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(sup1), 235–245.

Newman, J. H. C. (1852/1982). *The idea of a university*. University of Notre Dame Press.

O'Leary, M., Reynolds, K., Ling, G., Liu, O. L., Belton, S., O'Reilly, N., & McKenna, J. (2020). Assessing critical thinking in higher education: Validity evidence for the use of the HEIghten™ critical thinking test in ireland. *Journal of Higher Education Theory and Practice*, *20*(12), 115–130.

Orona, G. A. (2021a). Broken promises? Examining the effectiveness of promising practices in STEM lectures by student subgroups. *Innovative Higher Education*, *46*(2), 223–239.

Orona, G. A. (2021b). Gotta know why! Preliminary evidence supporting a theory of virtue learning as applied to intellectual curiosity. *Theory and Research in Education*, *19*(3), 279–295.

Orona, G. A. (2021c). Philosophy's rematch: A new conceptualization of the study of higher education. *Arts and Humanities in Higher Education*, *20*(4), 386–402.

Orona, G. A., Li, Q., McPartlan, P., Bartek, C., & Xu, D. (2022). What predicts the use of interaction-oriented pedagogies? The role of self-efficacy, motivation, and employment stability. *Computers & Education*, 104498.

Orona, G. A., & Pritchard, D. (2021). Inculcating curiosity: Pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment & Evaluation in Higher Education*, 1–15.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme.* New York: Holt, Rinehart; Winston.

Powell, C., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences*, *95*, 147–151.

Radovic, S., Firssova, O., Hummel, H., & Vermeulen, M. (2021). Improving academic performance in higher education: Strengthening the relation between theory and practice through prompted reflection. *Active Learning in Higher Education*.

Rhemtulla, M., Bork, R. van, & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30.

Rios, J. A., Guo, H., Mao, L., & Liu, O. L. (2017). Evaluating the impact of careless responding on aggregated-scores: To filter unmotivated examinees or not? *International Journal of Testing*, *17*(1), 74–104.

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, *49*(2), 80–89.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Schwartz, M. S., & Fischer, K. W. (2003). Building vs. Borrowing: The challenge of actively constructing ideas in post-secondary education. *Liberal Education*, *89*(3), 22–29.

Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., … others. (2020). Thinking critically about critical thinking: Validating the russian HEIghten critical thinking assessment. *Studies in Higher Education*, *45*(9), 1933–1948.

Sinnott, J. D. (1981). The theory of relativity: A metatheory for development? *Human Development*, *24*(5), 293–311.

Sinnott, J. D. (1998). *The development of logic in adulthood: Postformal thought and its applications*. Springer Science & Business Media.

Sinnott, J. D. (2008). Cognitive development as the dance of adaptive transformation: Neo-piagetian perspectives on adult cognitive development. In M. C. Smith & N. DeFrates-Densch (Eds.), *Handbook of research on adult learning and development* (pp. 125–156). Routledge.

Stanny, C., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education*, *40*(7), 898–913.

Van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., … others. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26.

Van de Schoot, R., Veen, D., Smeets, L., Winter, S. D., & Depaoli, S. (2020). A tutorial on using the WAMBS checklist to avoid the misuse of bayesian statistics. *Small Sample Size*

Solutions: A Guide for Applied Researchers and Practitioners; Van de Schoot, R., Miocevic, M., Eds, 30–49.

Vista, A. (2020). Data-driven identification of skills for the future: 21st-century skills for the 21st-century workforce. *Sage Open*, *10*(2), 2158244020915904.

Von Stumm, S. (2013). Investment traits and intelligence in adulthood. *Journal of Individual Differences*.

Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574–588.

Vu, V. Q. (2017). *Documenting instructional practices in large introductory STEM lecture courses* (PhD thesis). UC Irvine.

Wang, J.-S., Pascarella, E. T., Nelson Laird, T. F., & Ribera, A. K. (2015). How clear and organized classroom instruction and deep approaches to learning affect growth in critical thinking and need for cognition. *Studies in Higher Education*, *40*(10), 1786–1807.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*. Taylor & Francis.

Watson, M. W., & Fischer, K. W. (1977). A developmental sequence of agent use in late infancy. *Child Development*, 828–836.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS One*, *11*(3), e0152719.

Wineburg, S., Breakstone, J., McGrew, S., & Ortega, T. (2016). Evaluating information: The cornerstone of online civic literacy; executive summary. Stanford, CA: Stanford History Education Group.

Wolff, R., & Booth, M. (2017). Bridging the gap: Creating a new approach for assuring 21st century employability skills. *Change: The Magazine of Higher Learning*, *49*(6), 51–54.

Wysocki, A., Lawson, K. M., & Rhemtulla, M. (2020). Statistical control requires causal justification.

York, R. (2018). Control variables and causal inference: A question of balance. *International Journal of Social Research Methodology*, *21*(6), 675–684.

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*(4), 305–320.

Tables and Figures

Table 3.1
Descriptive Statistics

	N	M	SD
OuterBreadth	260	11.18	8.09
InnerBreadth	260	24.71	11.29
Hs_GPA	260	4.02	0.24
sat	260	0.33	0.81
d_pubhealth	260	0.07	-
d_bio	260	0.22	-
d_bus	260	0.03	-
d_soe	260	0.05	-
d_engi	260	0.13	-
d_hum	260	0.03	-
d_info	260	0.05	-
d_nurs	260	0.03	-
d_phys	260	0.05	-
d_seco	260	0.13	-
d_ss	260	0.15	-
d_art	260	0.01	-
d_male	260	0.30	-
d_junior	260	0.22	-
age	260	18.53	1.57
d_urm	260	0.28	-
d_full_time	260	1.00	-
sesIndex	260	0.01	0.93
nIC	260	0.54	0.21
openness	260	-0.06	1.00
NFC	260	3.50	0.56
EC	260	3.24	0.56
polisoc	260	3.30	0.86
nEffort	260	0.68	0.18
consci	260	-0.04	0.95
Optim	260	4.52	0.59
ASE	260	3.52	0.69
msf19nfge1	260	3.55	0.89
msf19jtsk2	260	3.83	0.82

Note. All variables beginning with "d_" indicate dummy coded variables where the focus group is coded 1 and the reference group 0. Thus, means can be interpreted as proportions. For the major schools listed: ss refers to social sciences; seco = social ecology, phys = physics, info = informatics and computer science; hum = humanities;

engi = engineering; soe = school of education; bus = business; bio = biology.

Table 3.2
Bayesian Structural Equation Model Fit Indices

	BRMSEA	BGamma Hat	adjBGamma Hat	BMc	BCFI	BTLI	BNFI	PPP
	0.077	0.97	0.936	0.941	0.914	0.872	0.869	0.007
M1: No CU's, no prior	(0.009)	(0.007)	(0.014)	(0.013)	(0.02)	(0.03)	(0.018)	
	0.05	0.988	0.968	0.975	0.964	0.938	0.923	0.172
M2: No CU's, prior	(0.021)	(0.008)	(0.021)	(0.016)	(0.024)	(0.042)	(0.023)	
	0.05	0.988	0.969	0.975	0.964	0.939	0.923	0.172
M3: 4 CU's, prior	(0.021)	(0.008)	(0.02)	(0.016)	(0.023)	(0.042)	(0.022)	
	0.05	0.988	0.969	0.975	0.964	0.939	0.923	0.202
M4: CT meta-factor, 2 CU's, prior	(0.021)	(0.008)	(0.02)	(0.016)	(0.023)	(0.042)	(0.022)	

Note. Posterior means for each fit index is shown; standard deviations are in parenthesis.

Table 3.3
Posterior Summary: Out of Major Course Breadth on Complex Cognitive Reasoning (CCR) Gain

	Posterior Mean	Posterior SD	Lower 89% Credible Interval	Higher 89% Credible Interval	Convergence Diagnostic	Effective Samples
Outer (Major) Breadth	0.17	0.08	0.05	0.29	1	3,655
CCR1	0.41	0.06	0.31	0.52	1	4,643
Intellectual Curiosity	0.02	0.07	-0.09	0.12	1	3,529
Effort	0.04	0.07	-0.07	0.16	1	3,755
Political & Social Involvement	0.11	0.06	0.01	0.21	1	4,486
Monetary Labor focus	-0.05	0.06	-0.15	0.04	1	5,456
Desire for GE	-0.08	0.06	-0.19	0.02	1	4,610
Age	0.02	0.05	-0.07	0.1	1	4,296
Male	0.1	0.14	-0.11	0.32	1	4,582
Urm	-0.29	0.14	-0.51	-0.07	1	4,625
SES Index	0.05	0.06	-0.05	0.16	1	5,304
Full-time	-0.13	0.67	-1.2	0.94	1	6,134
Junior	0.2	0.2	-0.12	0.53	1	4,230
HSGPA	0.07	0.28	-0.37	0.52	1	5,281
SAT	0.1	0.09	-0.04	0.23	1	4,525
d_pubhealth	-0.74	0.33	-1.25	-0.21	1	1,810
d_bio	-0.3	0.22	-0.66	0.06	1	1,363
d_bus	-0.01	0.35	-0.56	0.55	1	2,456
d_soe	-0.2	0.32	-0.7	0.29	1	1,871
d_engi	-0.15	0.25	-0.56	0.25	1	1,617
d_hum	0.12	0.33	-0.41	0.65	1	2,513
d_info	-0.18	0.31	-0.67	0.31	1	1,854
d_nurs	0.15	0.36	-0.42	0.73	1	2,893
d_phys	0.02	0.31	-0.49	0.51	1	2,042
d_seco	-0.18	0.24	-0.56	0.2	1	1,570
d_ss	-0.21	0.24	-0.59	0.17	1	1,370
d_art	-0.25	0.51	-1.03	0.59	1	4,426
sigma	0.88	0.04	0.82	0.95	1	6,131
(Intercept)	-0.26	1.66	-2.91	2.44	1	5,393

Note. Note. All variables beginning with "d_" indicate dummy coded variables where the focus group is coded 1 and the reference group 0. Thus, means can be interpreted as proportions. For the major schools listed: ss refers to social sciences; seco = social ecology, phys = physics, info = informatics and computer science; hum = humanities; engi = engineering; soe = school of education; bus = business; bio = biology. CCR1 = Complex cognitive reasoning at time 1.

Table 3.4
Prior Sensitivity Analysis:
Checking Different Prior Specifications

Priors	Posterior Mean	Posterior SD	Lower 89% Credible Interval	Higher 89% Credible Interval
Dogmatic Prior: $\beta_1 OuterBreadth \sim N(.13, .01)$	0.13	0.01	0.12	0.15
Informative 1: $\beta_1 OuterBreadth \sim N(.13, .08)$	0.15	0.06	0.06	0.24
Informative 2: $\beta_1 OuterBreadth \sim N(.13, .2)$	**0.17**	**0.08**	**0.05**	**0.29**
Informative 3: $\beta_1 OuterBreadth \sim N(.13, .5)$	0.17	0.08	0.05	0.3
Weakly Informative 1: $\beta_1 OuterBreadth \sim N(.13, 1)$	0.18	0.08	0.04	0.3
Weakly Informative 2: $\beta_1 OuterBreadth \sim N(.13, 4)$	0.18	0.08	0.04	0.31
Default: $\beta_1 OuterBreadth \sim N(0, 1)$	**0.20**	**0.09**	**0.06**	**0.33**
Diffuse/Flat: $\beta_1 OuterBreadth \sim N(0, 500)$	0.20	0.09	0.06	0.34

Note. The first bolded prior is the author's prior and was used to generate the interpreted posterior distributions. The second bolded prior is a default prior. Default priors refer to the priors automatically specified in the *stan_glm* function. These priors are meant to provide only weak or moderate regularization. For the default prior model, all parameters incorporated default priors.

Table 3.5
Posterior Summary for Individual Performance Assessments

	Informative Priors				Weak/Default Priors			
	Analytic	Synthetic	COR	PT	Analytic	Synthetic	COR	PT
Outer Breadth	0.152 (0.077)	0.152 (0.08)	0.099 (0.079)	0.081 (0.084)	0.186 (0.09)	0.162 (0.089)	0.105 (0.091)	0.077 (0.098)
Pretest	0.306 (0.063)	0.399 (0.066)	0.311 (0.06)	0.191 (0.068)	0.299 (0.063)	0.397 (0.063)	0.3 (0.062)	0.178 (0.069)
Demographics/Academic Variables/Attitudes	X	X	X	X	X	X	X	X

Note. Default priors refer to the priors automatically specified in the *stan_glm* function. These priors are meant to provide only weak or moderate regularization. For the default prior model, all parameters incorporated default prior: $\beta_1 OuterBreadth \sim N(0, 1)$.

Table 3.6
Robustness Checks

	Total Courses Taken					Within Major (Inner Breadth) Courses Taken					
	Posterior Mean	Posterior SD	Lower 89% Cred Interval	Higher 89% Cred Interval	Convergence Diagnostic		Posterior Mean	Posterior SD	Lower 89% Cred Interval	Higher 89% Cred Interval	Convergence Diagnostic
Total Courses Taken	0.044	0.099	-0.12	0.2	1	Within Major Courses Taken	-0.158	0.113	-0.338	0.026	1
d_pubhealth	-0.271	0.333	-0.795	0.281	1	d_pubhealth	-0.701	0.429	-1.383	-0.005	1
d_bio	-0.222	0.276	-0.669	0.209	1	d_bio	-0.301	0.284	-0.763	0.161	1
d_bus	0.168	0.434	-0.52	0.862	1	d_bus	-0.193	0.461	-0.921	0.547	1
d_soe	0.145	0.379	-0.455	0.75	1	d_soe	-0.19	0.42	-0.856	0.487	1
d_engi	-0.237	0.293	-0.72	0.221	1	d_engi	-0.121	0.308	-0.621	0.36	1
d_hum	0.348	0.41	-0.303	1.008	1	d_hum	0.106	0.408	-0.53	0.769	1
d_info	-0.071	0.409	-0.73	0.599	1	d_info	-0.501	0.417	-1.162	0.162	1
d_nurs	0.31	0.429	-0.377	1.012	1	d_nurs	0.101	0.458	-0.631	0.835	1
d_phys	0.213	0.353	-0.353	0.782	1	d_phys	0.1	0.362	-0.471	0.672	1
d_seco	-0.071	0.334	-0.603	0.466	1	d_seco	-0.376	0.339	-0.905	0.176	1
d_ss	-0.086	0.303	-0.573	0.387	1	d_ss	-0.296	0.313	-0.784	0.2	1
d_art	-0.004	0.609	-0.978	0.962	1	d_art	-0.257	0.627	-1.255	0.733	1
CCR1	0.419	0.06	0.323	0.514	1	CCR1	0.423	0.061	0.328	0.521	1
NFC	0.103	0.112	-0.073	0.281	1	NFC	0.128	0.114	-0.056	0.307	1
consci	-0.005	0.067	-0.112	0.102	1	consci	-0.007	0.065	-0.113	0.097	1
age	-0.013	0.054	-0.098	0.075	1	age	0	0.056	-0.09	0.09	1
GenderM	0.083	0.138	-0.136	0.303	1	GenderM	0.091	0.134	-0.123	0.306	1
Urm	-0.277	0.135	-0.493	-0.064	1	Urm	-0.281	0.136	-0.499	-0.069	1
sesIndex	0.035	0.066	-0.07	0.139	1	sesIndex	0.031	0.065	-0.074	0.134	1
statusjunior	0.149	0.209	-0.195	0.475	1	statusjunior	0.165	0.206	-0.159	0.491	1
Hs_GPA	0.135	0.291	-0.333	0.603	1	Hs_GPA	0.149	0.286	-0.298	0.61	1
sat	0.094	0.088	-0.047	0.234	1	sat	0.099	0.089	-0.041	0.245	1
(Intercept)	-0.596	1.623	-3.161	1.971	1	(Intercept)	-0.818	1.622	-3.448	1.704	1
sigma	0.895	0.041	0.832	0.964	1	sigma	0.892	0.04	0.829	0.957	1

Note. Note. All variables beginning with "d_" indicate dummy coded variables where the focus group is coded 1 and the reference group 0. Thus, means can be interpreted as proportions. For the major schools listed: ss refers to social sciences; seco = social ecology, phys = physics, info = informatics and computer science; hum = humanities; engi = engineering; soe = school of education; bus = business; bio = biology. CCR1 = Complex cognitive reasoning at time 1. CCR1 = complex cognitive reasoning at time 1.

Table 3.7
Model Comparison

	Posterior Mean (β_1)	Posterior SD	Lower 89% Credible Interval	Higher 89% Credible Interval	WAIC	Effective # of Parameters	Bayesian R^2
DAG Model	**0.154**	**0.076**	**0.031**	**0.276**	**700.343**	**22.535**	**0.304**
Bivariate	0.069	0.061	-0.031	0.165	741.503	2.743	0.303
Pretest	0.075	0.052	-0.01	0.158	741.503	2.743	0.223
Pretest + Academic	0.112	0.067	0.006	0.22	690.678	14.926	0.277
Pretest + Academic + Demographics	0.145	0.077	0.023	0.265	696.932	19.921	0.297
Full Model (+ Attitudes/Dispositions)	**0.167**	**0.078**	**0.043**	**0.293**	**699.514**	**24.837**	**0.320**

Note. WAIC = Watanabe-Akaike information criterion. DAG Model includes predictors that close the backdoor paths presented in figure 1. These include: CCR1, age, attitude/disposition variables, major, extra-curricular honor's program, and prior academic achievement. β_1 = Outer breadth parameter.

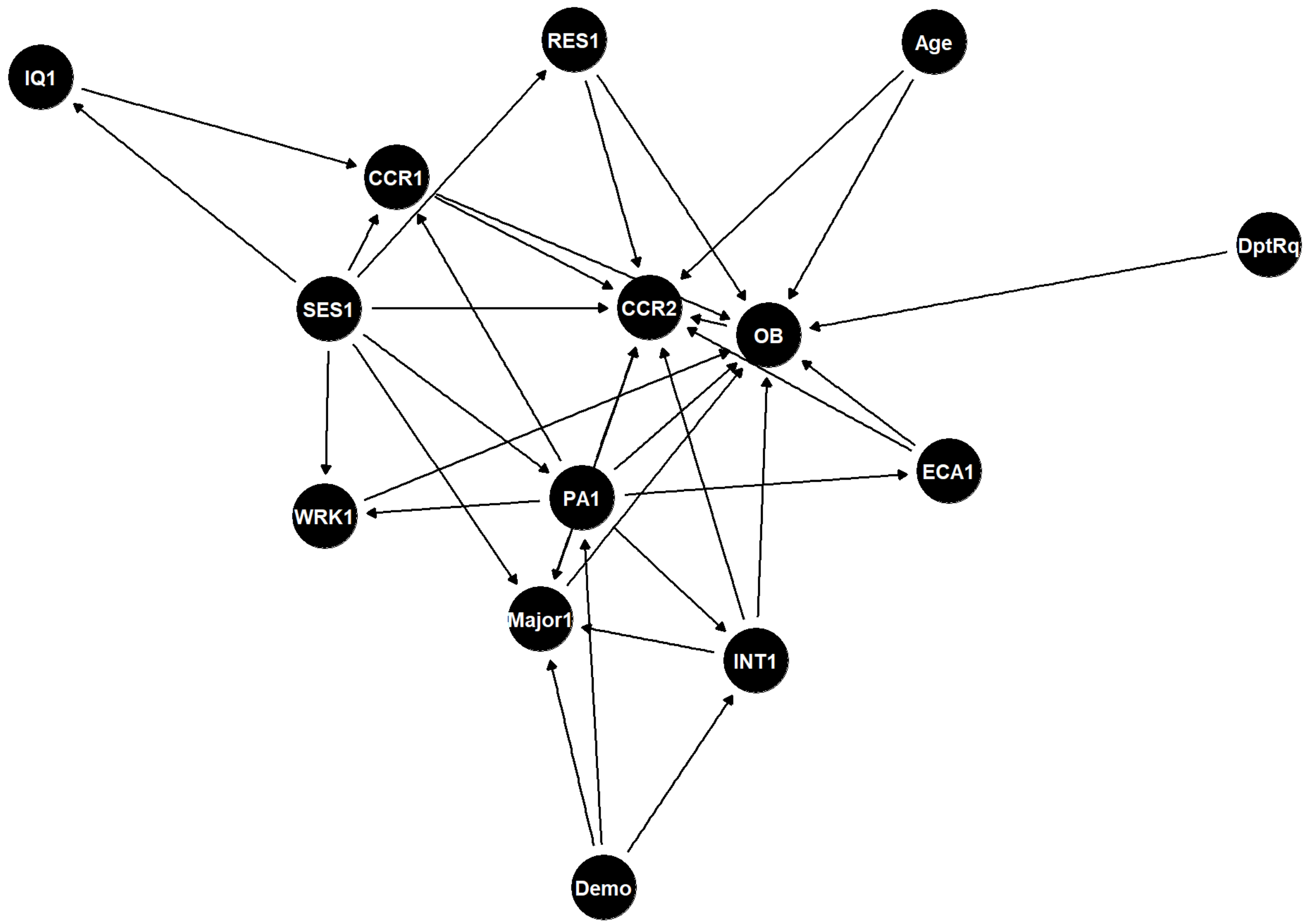


Figure 3.1. Directed acyclic graph (DAG) depicting the hypothesized causal structure. SES1 = socioeconomic status; PA = prior academic achievement; INT1 = dispositions/attitudes; OB = outer breadth; CCR = complex cognitive reasoning; WRK1 = whether a student works; ECA1 = extracurricular activities; Demo = demographics; DptRq = department requirements; RES1 = on-campus residence.

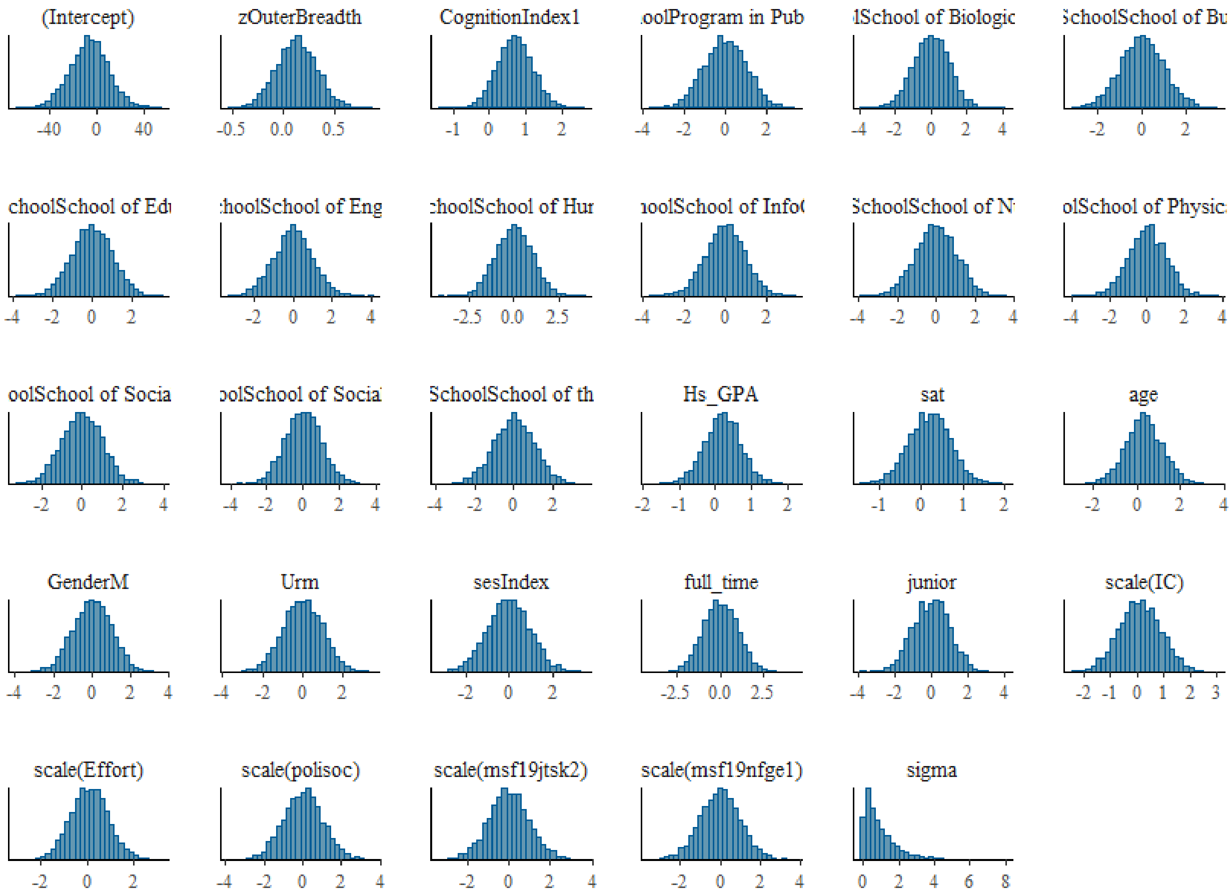


Figure 3.2. Prior predictive checking: simulations for every parameter.

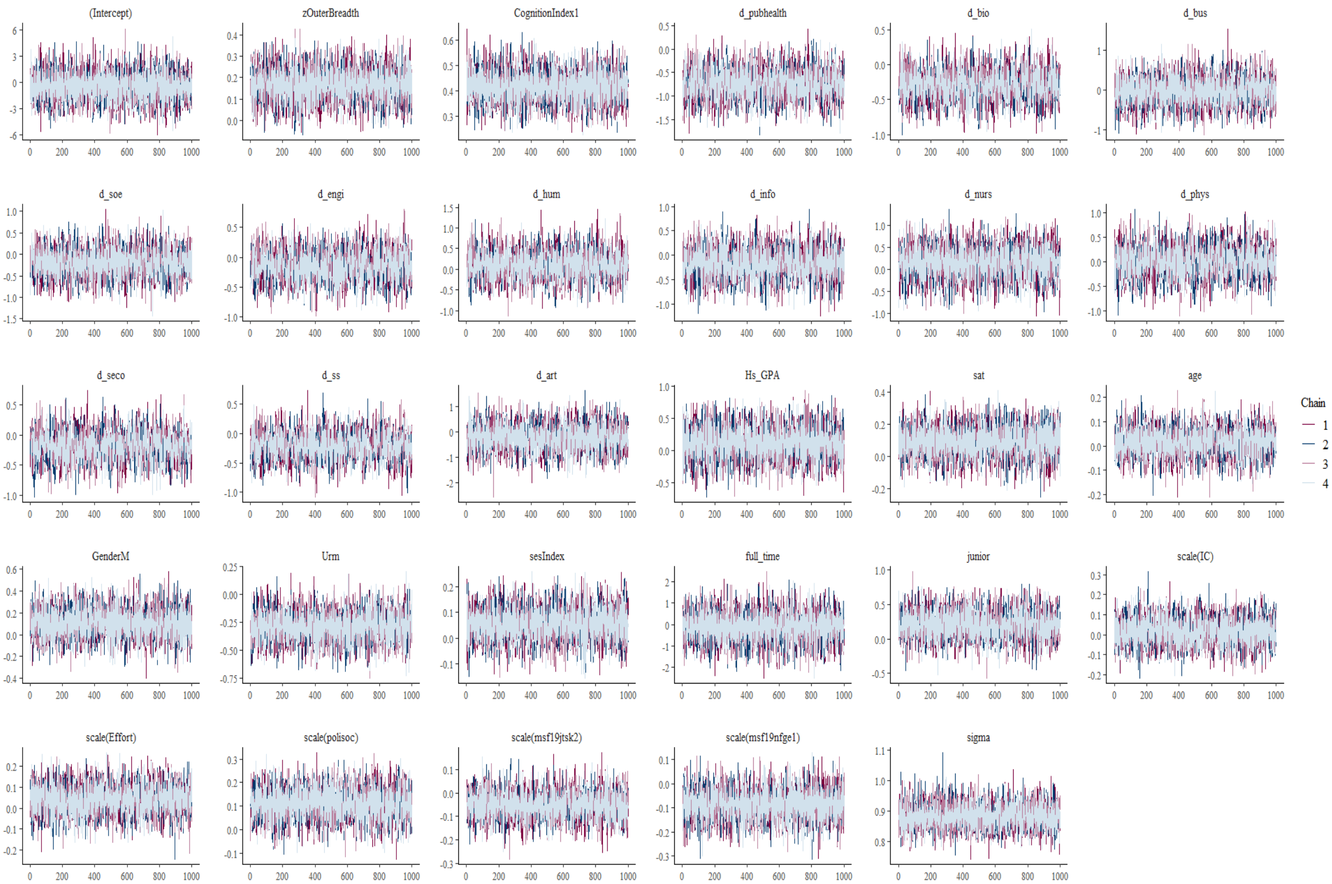


Figure 3.3. Trace plots to check convergence for each parameter in the model.

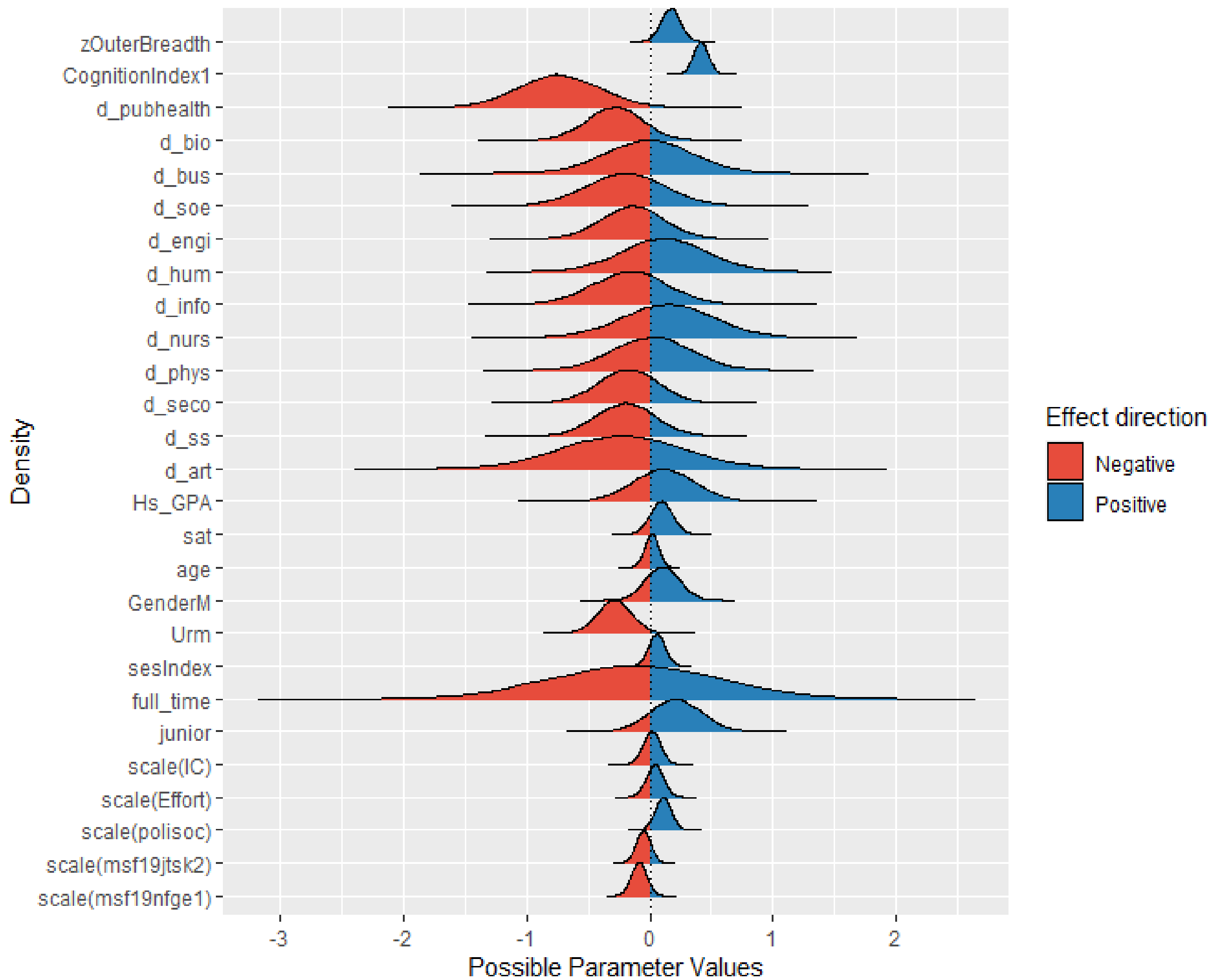


Figure 3.4. Marginal posterior distributions for each parameter in the model.

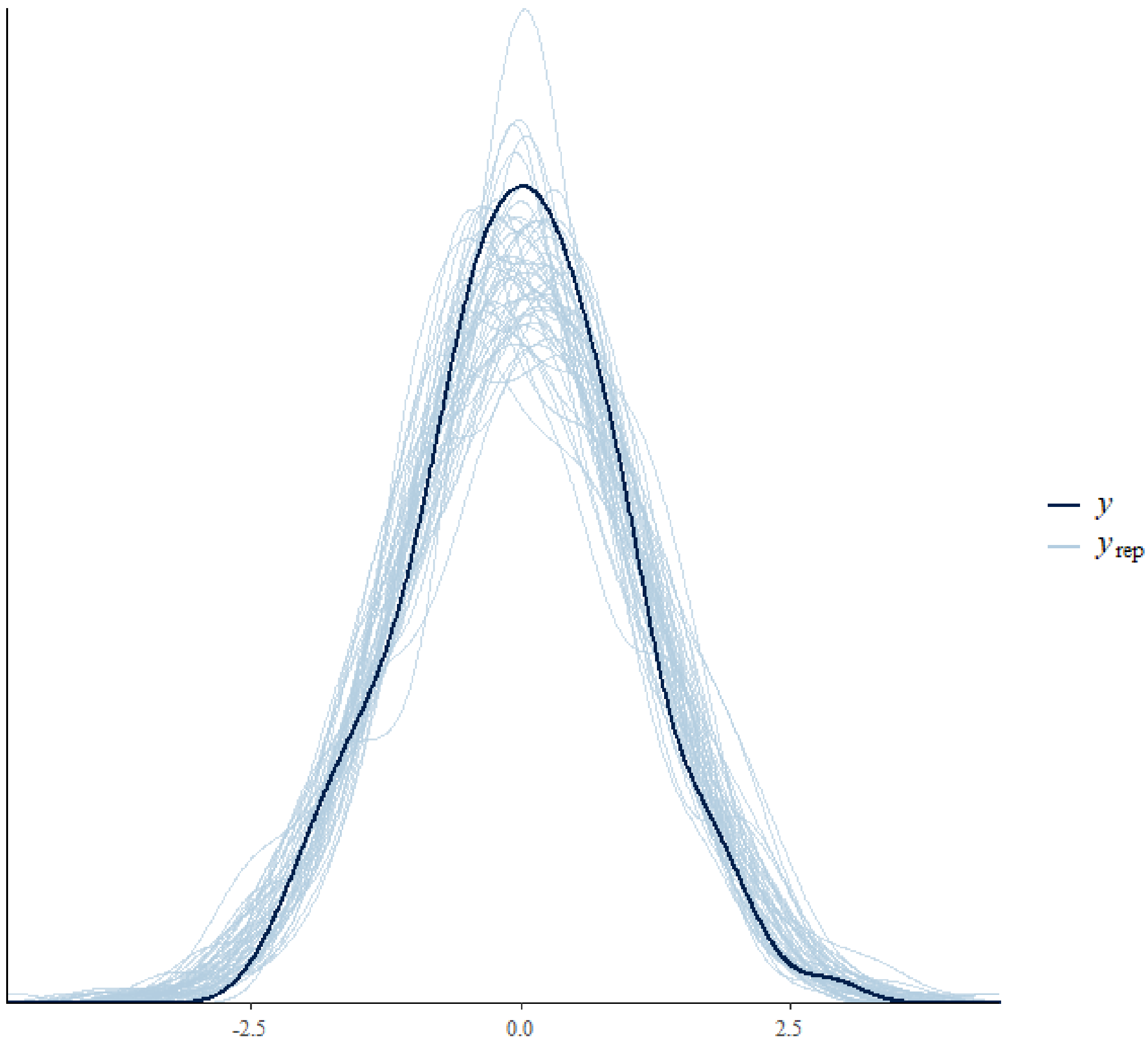


Figure 3.5. Posterior predictive check.

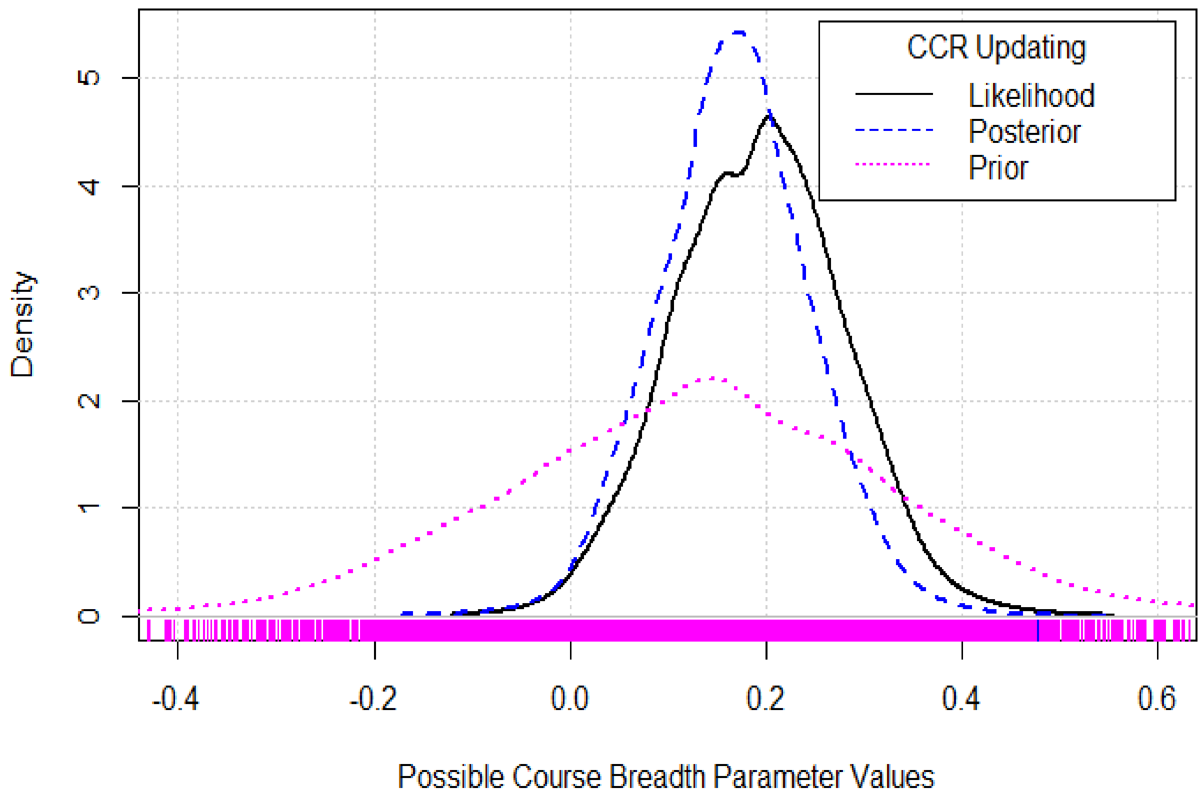


Figure 3.6. Bayesian updating showing the prior, posterior and likelihood distribution of the of the regression estimate for course-taking breadth predicting complex cognitive reasoning at time 2, controlling for CCR1, attitudes/dispositions, demographics, and academic variables. It should be noted that the dark black line is proportional to the likelihood, but not the likelihood strictly speaking.

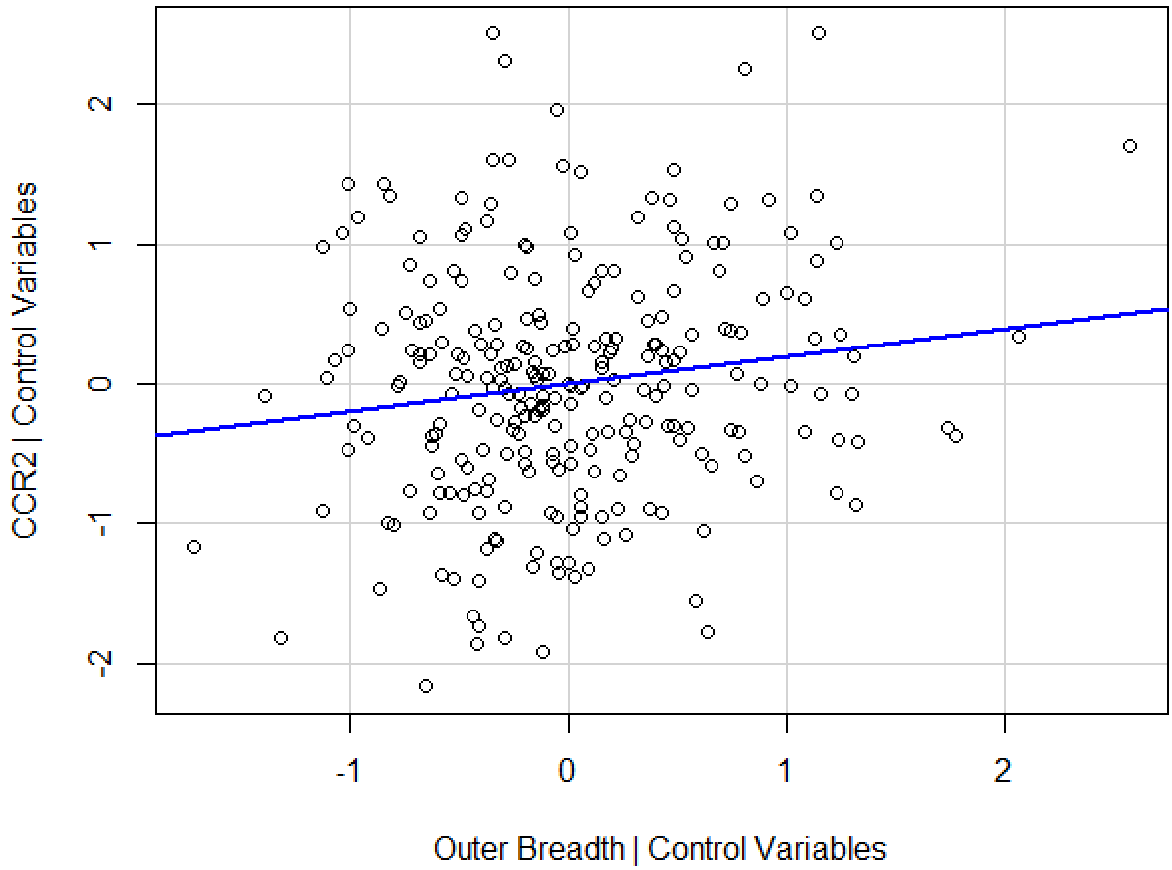


Figure 3.7. Partial regression plot depicting the relationship between outer breadth and CCR2 (Complex cognitive reasoning at time 2), controlling for all other variables in the model.

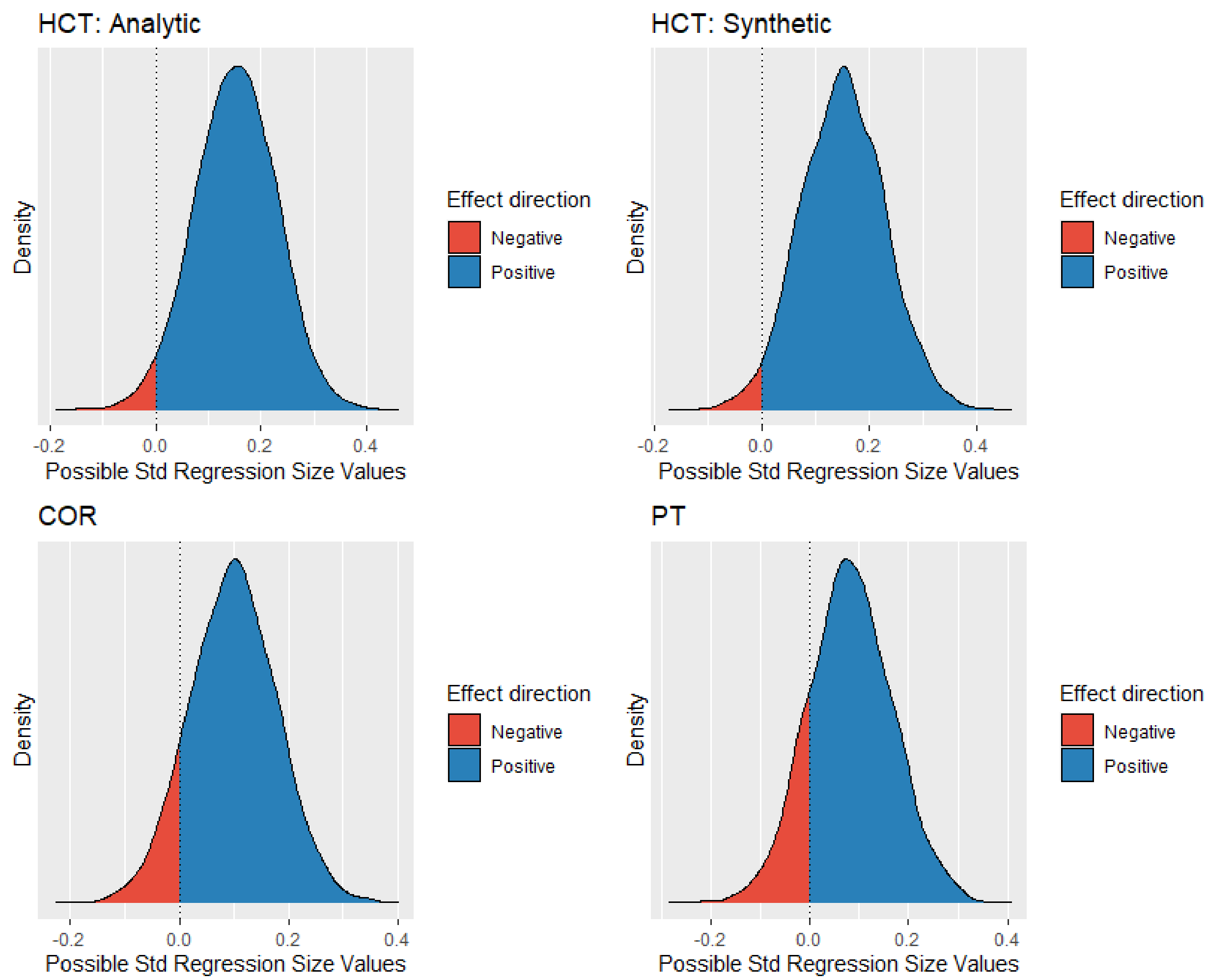


Figure 3.8. Marginal posterior distribution for the relationship between outer breadth on each individual task constituting the complex cognitive reasoning skill domain. HCT = Heighten Critical Thinking; COR = Civic Online Reasoning; PT = Perspective-Taking.

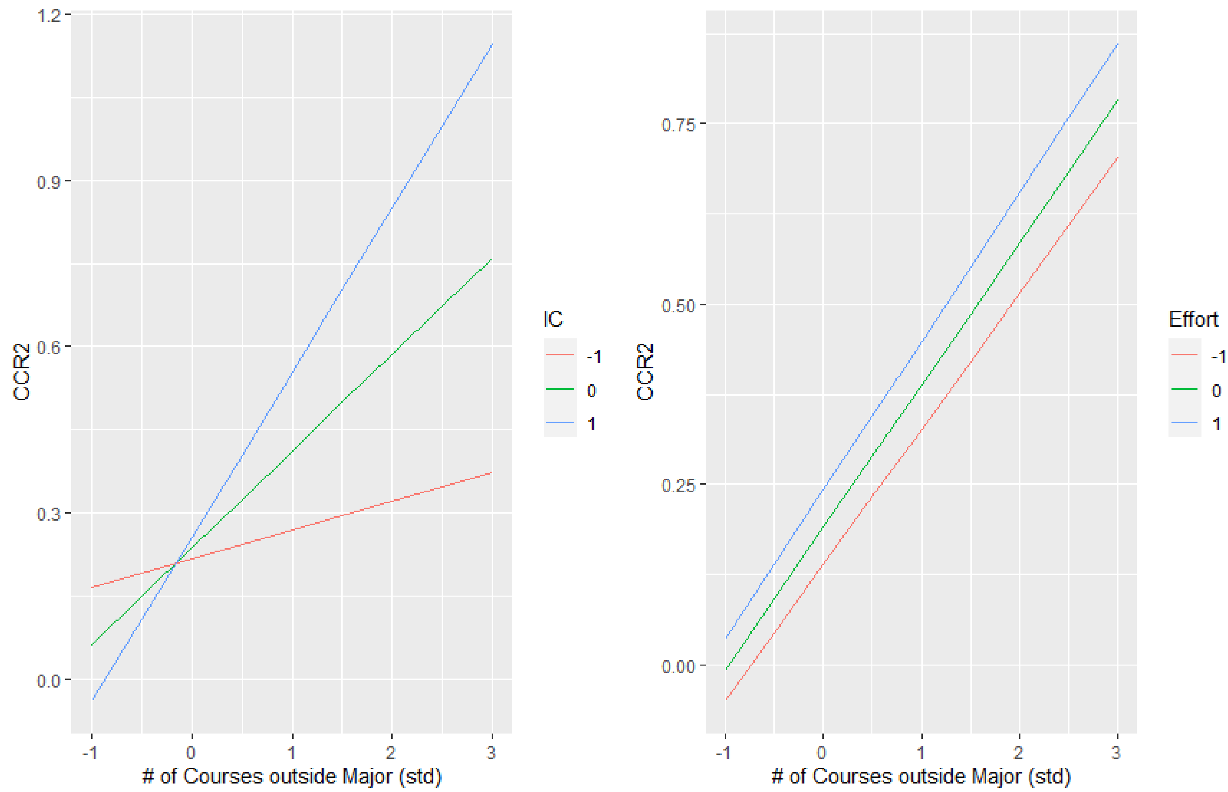


Figure 3.9. Interaction graphs. The first pane is IC (grouped by standard deviations) interacted with outer breadth and the second is effort interacted with outer breadth.

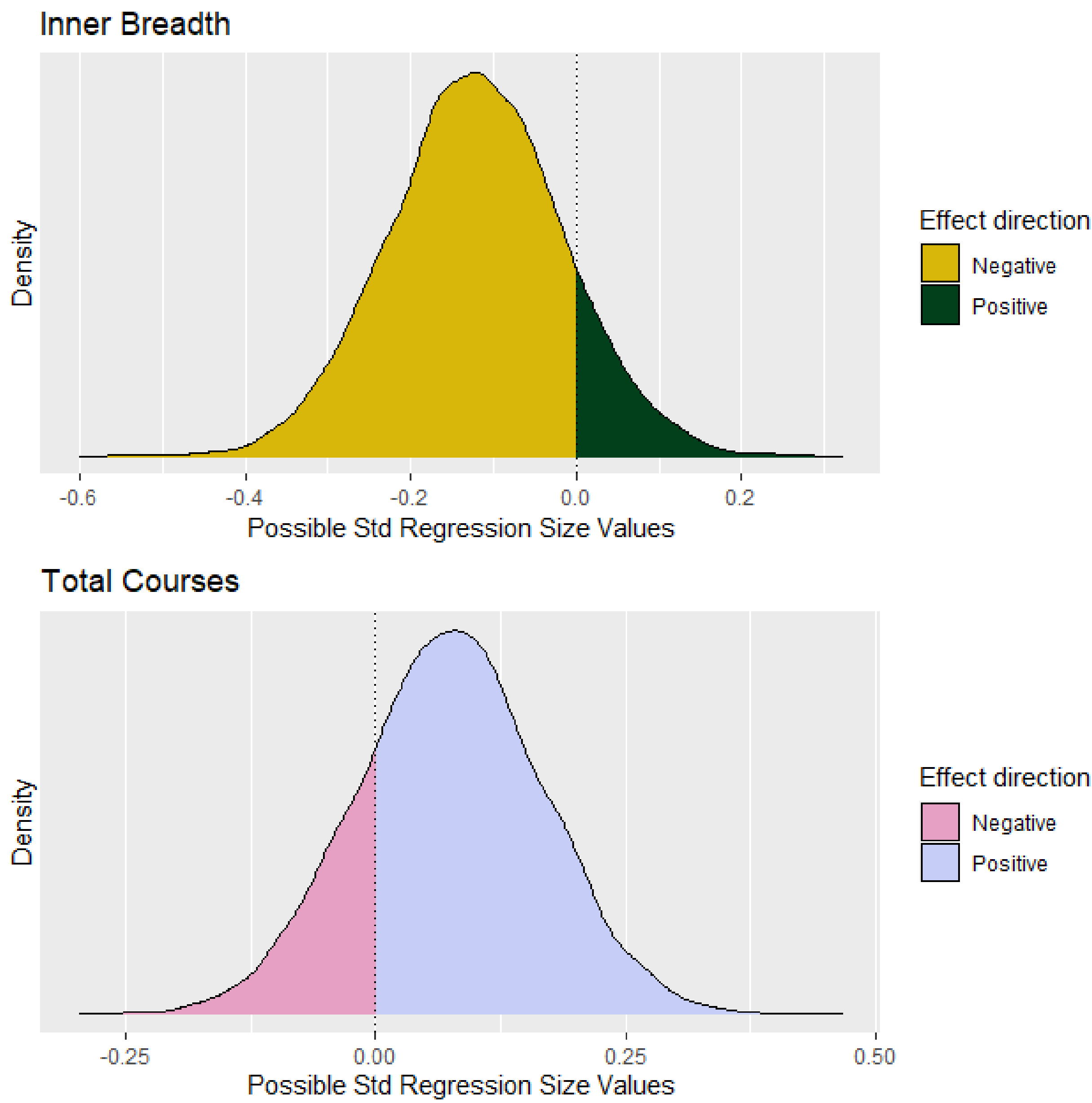


Figure 3.10. Marginal posterior distribution for the relationship between inner breadth and total courses on complex cognitive reasoning.

CHAPTER 4

Summary and Conclusions

Overall Summary

Each of the three studies presented above attempted to understand undergraduate experiences that lead to liberal art outcomes associated with cognition and character. The studies aimed to glean information regarding if and how practices are beneficial to students. More specifically, the studies examined if a classroom-based intervention could bolster intellectual curiosity within courses, tested the possible mechanisms for curiosity development, and examined the extent to which general education is associated with gains in complex reasoning skills.

I found that that the intervention, designed based upon theories of virtue epistemology and interest theory, showed preliminary effectiveness in inculcating curiosity. I also found that, as predicted by a general theory of virtue learning, learning what intellectual virtue is and why it's important are likely important determinants to developing this characteristic. Finally, motivated as a supportive context according to skill theory, exposure to diverse sources of educational content appear to be associated with the development of cognitive abilities, though it's unclear whether the magnitude of these associations warrant the ubiquity of general education requirements.

Moreover, I found equal satisfaction with the intervention across important student subgroups, such as ethnicity, socio-economic status, first-generation status, and sex. Other findings include the importance of intellectual curiosity, which was shown to moderate the influence of out-of-major course taking breadth on the development of complex cognitive reasoning. And finally, I found that out-of-major course taking breadth displayed a weaker relationship with ill-structured tasks (e.g., perspective-taking) than well-structured (e.g., argument evaluation).

Each of these studies are independent and thus distinct. However, together they form a coherent set regarding if and how college-going experiences relate to the development of soft and hard cognitive skills—i.e., cognition and character. Furthermore, each study has clear implications for future research. I review these possibilities below.

Future Directions

Study 1. Study 1 introduced a novel online module to inculcate intellectual virtue, beginning with the foundational virtue of curiosity. Given that the results showed preliminary effectiveness across several outcomes, three clear next steps are presented: (a) build out, measure, and assess more intellectual virtues, (b) administer cognitive performance assessments to both validate new intellectual virtue scales and examine the effect of the module on more objective measures, and (c) implement a randomized control trial (RCT) to estimate the unbiased causal impact of the intellectual virtue curriculum.

This expansion would allow for a much more robust evaluation. Assessing more intellectual virtues, such as intellectual humility, integrity, and tenacity, itself can open many new and important research pathways. For instance, with these data, which of the virtues is most impacted by the module and which virtue is most correlated with student grades and performance would be answerable research questions. Moreover, it would allow for overall summary evaluations and psychometric hypothesis testing regarding the presence of meta-factors.

Implementing cognitive performance assessments is a vital next step as issues with self-report scales are well-documented (Maul, 2017), and should not be relied upon as the sole outcome of evaluation studies. However, any objective measure will not be suitable: it must have at minimum two features: (a) a cognitive ability component and (b) a character component. Most critical thinking assessments mostly emphasize the former attribute and tend to be highly correlated with SAT scores and other intelligence tests. A promising assessment that could be used for this purpose is the Cognitive Reflection Test (Frederick, 2005), with lower scores representing cognitive miserliness and higher one's engagement. Originating in the economics literature, the test has been expanded and implemented in a wide variety of settings by researchers attempting to understand the intersection of ability and thinking styles (Toplak et al., 2011).

Finally, an RCT would greatly strengthen our understanding of the causal effect of the intellectual virtue curriculum. Such a design would boost our confidence in the efficacy of the modules and, with the concomitant expansion of new measures, would enable gaining deeper knowledge of how the intervention impacts students. For instance, heterogeneity across subgroups could be tested across each intellectual virtue.

Study 2. Study 2 tested the mechanisms by which growth in intellectual curiosity occurs. The results tested explicit predictions from Besser's theory of learning virtue. The viable next steps for this line of research parallel that of study 1: (a) expand the measures used to measure the theoretical entities in Besser's theory, (b) test the theory across multiple intellectual virtues, and (c) embed the intervening analysis within a RCT design.

A weakness of study 2 was the use of one item each for the two elements causing virtue growth in Besser's theoretical model, with the third construct (knowing *how* to exercise virtue) going unmeasured. A future study could feasibly generate more items for all the theoretical constructs in Besser's theory, test their reliability and validity, and subsequently specify a structural equation model to account for measurement error when making predictions.

Study 2 only examined intellectual curiosity as an outcome. Like study 1, a future study could build out more intellectual virtue measures to test the applicability of Besser's theory across different virtue types. This information could inform specific revisions to the theory. The data could also be used to test competing models, each being instantiations of theory.

Finally, an RCT could be implemented to test the causal link of an intervention that impacts students' knowledge and value for intellectual virtue, which in turn impacts their actual level of intellectual virtue. In study 1, the module was shown to have a preliminary association with both intellectual curiosity and the two elements in Besser's theory; and study 2 highlighted how these Besser elements are correlated with intellectual curiosity. Therefore, combining these three pathways (X→Z→Y and X→Y) in one model via mediation is a plausible next step.

Study 3. Study 3 examined the relationship between course-taking breadth and the development of complex cognitive reasoning abilities. This study has many avenues for future research. These include: (a) a cost-benefit analysis of out-of-course breadth, (b) more fine-grained analysis of course content to generate intellectual breadth measures, (c) experimental design randomizing students to course breadth, and (d) exploratory work examining the precise educational combinations that are associated with complex reasoning gain.

Most of these future lines of research can be expounded upon at length. A cost-benefit analysis is necessary if general education and out-of-major breadth warrants continued funding. Such a study could systematically gather data on the cost of implementing and providing these course requirements and compare that with the dollar-value of corresponding gains in complex cognitive reasoning. Such a study may require an interdisciplinary team of researchers consisting of at minimum labor economists and cognitive psychologists, as well as higher education researchers.

For the second proposal, future research could possibly construct different measures of course-breadth that can be tested against the one used in study 3. Such measures might look at course-taking breadth across departments or the schools they are nested within and choose to include or exclude those that are similar to a student's major.

While typical college-going experiences allow for any student to take any course (conditional on prerequisites and availability), a feasible intervention could be constructed to approximate causal inference with more confidence. For instance, one feasible design could be to recruit students to participate in a novel educational experience where the coursework is largely determined for them (with some individual freedom). Then, one group could be randomized to a condition that exposes students to diverse ideas and general education via coursework (considering each individual major), and the other could be assigned to a condition where the coursework is limited to a small subset of courses related to the students major. The two groups could subsequently be compared on reasoning gains.

Finally, out-of-major breadth is vague. What courses matter, or rather, what combinations of courses matter? Future research could investigate the patterns of course-taking breadth that are associated with cognitive gain. For instance, perhaps it's necessary to have at least one course in a key field (STEM, social science, arts, and humanities, etc.). Or perhaps many unique combinations have similar effects, such as a lot of physics paired with humanities being equivalent to a lot of chemistry paired with social science. Such questions are highly relevant to maximizing the potential of higher education in affecting the cognitive development of its students.