



Epistemic virtue in higher education: testing the mechanisms of intellectual character development

Gabe Avakian Orona¹ · Duncan Pritchard² · Richard Arum² · Jacqueline Eccles² · Quoc-Viet Dang² · David Copp² · Daniel Alexander Herrmann² · Bruce Rushing² · Steffen Zitzmann¹

Accepted: 17 July 2023 / Published online: 26 July 2023
© The Author(s) 2023

Abstract

Epistemic virtues are character traits conducive to principled ways of thinking, leading to a life of flourishing. Recent years have witnessed an emergence of theoretical accounts describing how they develop. However, few if any studies have conducted rigorous empirical investigation into the mechanisms of intellectual virtue development. In this study, we review several significant frameworks before utilizing a randomized, pretest/posttest control trial design to understand the impacts of a novel thinking disposition intervention on intellectual virtue growth. We ascertain the direct and indirect effects of the intervention on four key intellectual virtues: curiosity, humility, integrity, and tenacity. Additionally, we test theoretical mediators of virtue learning. Clear evidence favoring a theory-inspired mediator is observed, though we observe weaker signals of direct effects, with nuances across the virtues. For instance, tenacity and curiosity variables appear to respond more to the intervention than do humility and integrity. Findings are discussed in light of contemporary theoretical perspectives.

Keywords Virtue science · Epistemic virtue · Intellectual virtue · Thinking disposition · Character education

Introduction

Intellectual character education, which refers to educating a person in a manner that inculcates desirable thinking dispositions known as epistemic or intellectual virtues—is gaining popularity in all sectors of education (Baehr, 2021; Ruch et al., 2020; Zhang et al., 2022), and particularly higher education (Arum, Eccles, Heckhausen et al., 2021; Orona, 2021c; Schwartz, 2020). The relevance is fueled, in part, by the growing sub-field of character education in psychology, the desire of employers to see graduates with specific soft skills inclusive of reflective thinking habits (McGrew et al., 2018; Villacís et al., 2022), and the widespread aim of educators to assist students in personal growth trajectories leading to a life of flourishing (Arum, Eccles et al., 2021).

While its significance is gaining steam across policy recommendations and scholarship, few empirical studies—and even less experimental—address whether educational processes produce such epistemic virtues, and whether specific interventions can be devised that espouse them (Ruch et al., 2020). In this study, we begin by succinctly discussing three key theoretical perspectives. These include: *virtue epistemology*, the *STRIVE-4 framework*, and *Besser’s virtue*

✉ Gabe Avakian Orona
gabriel.orona@uni-tuebingen.de

Duncan Pritchard
dhpritch@uci.edu

Richard Arum
richard.arum@uci.edu

Jacqueline Eccles
jseccles@uci.edu

Quoc-Viet Dang
qpdang@uci.edu

David Copp
dcopp@uci.edu

Daniel Alexander Herrmann
daherrma@uci.edu

Bruce Rushing
brushing@uci.edu

Steffen Zitzmann
steffen.zitzmann@uni-tuebingen.de

¹ Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany

² University of California, Irvine, US

learning theory. We highlight the connections between each and focus on noteworthy prior research before analyzing experimental data pertaining to several thinking disposition outcomes. Afterward, we test a theoretically inspired mediator as the possible link between the intervention and virtue growth. Finally, we discuss implications and future work.

Theoretical background

Virtue epistemology and intellectual virtue

The notion of an intellectual virtue can be traced to ancient Greece, being explicitly introduced in Aristotle's *Nicomachean Ethics* as a set of personal qualities and knowledge contributing to *eudaimonia*, or individual flourishing. Aristotle described two parts of the soul: one irrational and the other rational, with the latter separating humans from animals and characterizing the capacity for the development of moral and intellectual virtue. The two sets of virtue—moral and intellectual—are theorized to develop via different routes. While the moral virtues such as courage, generosity, temperance, and justice, to name a few, are developed by habituation and training from youth, the intellectual virtues are 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 *Nicomachean Ethics*, the best life is asserted to be the life of contemplation; a life marked by theoretical wisdom (Zhang et al., 2022).

The Aristotelian concept of intellectual virtue has since undergone modifications and has been used to resolve debates in the field of epistemology (Sosa, 1980), leading to the sub-field of virtue epistemology (Kotsonis, 2020), which evaluates processes of knowing and knowledge as stemming from these individual attributes. The most popular virtue epistemology thesis in recent decades is Zagzebski's (1996) "*Virtues of the Mind*." Zagzebski (1996) argues that intellectual virtues are just like moral virtues: attributes of a person that dictate good/desirable behavior (moral virtue); however, in the case of intellectual virtue, good/desirable thinking (Zagzebski, 1996). In this way, intellectual virtues ought to assist in the acquisition of true and accurate beliefs via good epistemic habits, hence they are also referred to as *epistemic virtues* (Greene & Yu, 2016).

Over the years, scholars have unsurprisingly described intellectual virtues differently. Whereas some define them as reliable cognitive faculties (Greco, 2000; Sosa, 1985), such as memory and perceptual acuity, Zagzebski (1996) promotes a twofold construct, with one component relating to

cognitive skills or capability, and the other component motivational. Baehr (2015) expanded on Zagzebski's model, adding two more dimensions: an affective and judgment dimension. The affective dimension goes beyond intellectual interest in a topic, or even motivation that leads to truth-seeking behavior, and emphasizes the *enjoyment of the learning process* itself. That is, one is intellectually virtuous not only if they desire to—and actually—solve an epistemic problem, but take pleasure in asking questions and being inquisitive. The judgement dimension adds yet another layer, which entails being sensitive to or "judging" when a particular moment or situation (in everyday life or otherwise) calls for critical reflection.

STRIVE-4 framework

Scholarship in moral virtue also offers insights into how virtue develops and what types of research questions are well-suited for virtue inquiry. Because of the growing interest in (intellectual) virtues, scholarship is beginning to transition from philosophical, psychological, and policy ruminations to a testable empirical framework (Fowers et al., 2021). For example, Fowers et al. (2021) introduced the STRIVE-4 (Scalar Traits that are Role sensitive, include Situation × Trait Interactions, and are related to important Values that help to constitute Eudaimonia) model to support and integrate the currently disjointed yet budding field of virtue science. The STRIVE-4 model posits that virtue has four components: knowledge, behavior, motivation, and disposition.

As the acronym suggests, virtues are explicitly conceptualized as quantitative attributes (e.g., individuals possess certain amounts of virtue) as opposed to categorical (e.g., presence/absence of virtue). Fowers et al., (2021, p. 129) also note "...from a neo-Aristotelian perspective, virtue traits are not biologically given. Rather, we see virtues as acquired traits." Though they are conceptualized as stable dispositions, virtues are role sensitive and dependent on specific contexts and situations.

With this basis, Fowers et al. (2021) generate 26 testable hypotheses stemming from their model. The hypotheses range from topics regarding measurement validity (e.g., #17. Self-reported virtue will be related to relevant criterion variables), anticipated correlations (e.g., #26. Virtue traits will be associated with variations in neural processes), prediction (#19. Virtue-related knowledge will add incremental validity to the ability to predict virtue-related criteria), development (e.g., #22. The rudiments of virtue can develop over time into mature virtue), and intervention (e.g., #24. Virtue-related behavior can be increased with simple, short-term interventions). The hypotheses are intended to guide virtue science, thus the STRIVE-4 framework does not

provide a list of virtues nor does it suggest specific mechanisms for development.

Besser's theory of virtue learning

Another noteworthy perspective in moral virtue is Besser's (2020) theory of virtue learning, which is grounded in self-determination theory. The theory notes that for virtue to develop in individuals, it must resonate with their perception of who they are and who they desire to become. Thus, it must connect with and become a part of their identity. In this way, the successful development of virtue must tap an individual's basic psychological needs, such as autonomy, competence, and relatedness.

In order for instruction to support virtue development by connecting it to learners' sense of autonomy, competence, and how they relate to others, the theory predicts that learning *what* virtue is/what is involved in its application, *why* it's important, and *how* to implement it in one's own life are the three key knowledge areas necessary for virtue learning. Besser (2020) places particular salience on learning *why* (e.g., why virtue is important), arguing that this is the bedrock for which deeper learning of the *what* and *how* of virtue follow suit. She (2020, p. 285) summarizes, "...our focus within moral development ought to be on helping subjects develop an understanding of the goal of virtue in a way that resonates with them."

Connections between the models

The frameworks and theories share important features with each other. First, all three agree that the aim of virtue leads to eudonmiona or flourishing. Second, they suggest that instruction and intervention are effective means of developing virtue. Third, two of the three frameworks explicitly mention that relaying the significance of virtue is a salient component of virtue instruction.

While both the STRIVE-4 model and Besser's theory were originally formulated to pertain to moral virtue, their scope readily extends to intellectual virtues. This is because: (a) the STRIVE-4 virtue components conform with proposed intellectual virtue dimensions (Baehr, 2015); (b) empirical studies of character often classify intellectual and moral virtue studies together (Brown et al., 2022; McGrath et al., 2020); (c) growth trajectories of intellectual and moral dispositions occur synchronously (King et al., 1989); (d) previous research has applied moral virtue theory to intellectual virtue testing (Orona, 2021b); (e) teaching strategies to inculcate moral and intellectual virtue have substantial overlap (Besser, 2020; Kotzee et al., 2019); (f) leading virtue taxonomies categorize several popular thinking dispositions as virtues (e.g., VIA-model listing curiosity and

open-mindedness as virtues) and (g) seminal philosophical texts argue against this dichotomization of virtue types (Zagzebski, 1996). Thus, the intellectual virtues appear to be sufficiently alike the moral virtues to warrant the application of the STRIVE-4 framework and Besser's theory to intellectual virtue research (Zagzebski, 1996).

(Intellectual) virtue research

Several empirical studies support aspects of the models reviewed above. For instance, one persistent issue is if (intellectual) virtue measures have appropriate (e.g., positive, negative or nonoverlapping) associations with external criteria, over and above other, related variables that have been studied for decades. Importantly, McGrath et al. (2020) found that broad character traits measured with the Virtues in Action (VIA) scale are highly related to, but not identical with, long-standing personality scales such as the NEO and HEXACO. Anjum and Amjad (2021) established this structure in a different population, confirming key associations with positive and negative affect.

Importantly, Lian and You (2017) found that several key virtues predict behaviors in undergraduates, such as time spent on smartphones. Self-reported measures of intellectual humility have been positively correlated with cognitive reflection (Krumrei-Mancuso, Haggard, LaBouff, & Rowatt, 2020) and mastery behaviors (Porter, Schumann, Selmeczy, & Trzesniewski, 2020). Orona et al. (2023) found that an index composed of various curiosity trait measures (e.g., Openness to Experience, Need for Cognition, and Epistemic Curiosity) moderated the influence of broad learning experiences on the development of higher-order cognition, such that highly curious individuals exhibited greater gains in reasoning ability when exposed to diverse educational content than their less curious peers. And one of the mainstay thinking dispositions—the Need for Cognition—has been positively related to cognitive reflection (Šrol, 2018) and significantly predicted complex problem-solving abilities (Rudolph et al., 2018).

Moreover, general character education (including intellectual character) programs have had positive effects on well-being and other desirable outcomes. Brown et al. (2022) found in a meta-analysis that most character education programs—across a range of intervention and outcome types—exhibit an average standardized effect size of 0.24. Interestingly however, few studies have looked at the effects of character education on character, and even less on intellectual character (Ruch et al., 2020).

The intellectual virtue curriculum development and prior research

Our research team developed and piloted a novel, online thinking disposition intervention internally referred to as the “Intellectual Virtue Curriculum” (IVC). The intervention was informed by philosophical, psychological, and online pedagogical theory and practice (Baehr, 2013; Fischer et al., 2022; Hidi & Renninger, 2006; Orona, Li et al., 2022), and the brainchild of a leading virtue epistemologist. In designing the IVC, theoretical perspectives were combined to trigger and maintain interest (Hidi & Renninger, 2006) in intellectual virtues via pedagogical activities, such as: (a) *introducing novel experiences* (Quinlan, 2019), (b) *being exposed to experts’ struggles and applications of concepts* (Hong & Lin-Siegler, 2012), (c) *interactive learning activities alongside a lecture course* (Yuretich et al., 2001), (d) *repeated involvement in inquiry activities*, and (e) *engaging in reflective exercises* (Hulleman & Harackiewicz, 2009).

Corresponding with the above theory-based pedagogical activities, the IVC consists of a set of modules containing high-quality videos of philosophers, scientists, educators, and graduate students detailing the components of intellectual virtue, their relevance, and the ways they relate to academic and intellectual pursuits, as well as a life of flourishing. Exemplars model good epistemic thinking across different domains and perspectives, an important feature in developing students’ epistemic thought processes (Vossoughi et al., 2021). Alongside videos, students are given a series of activities: including quizzes, thought puzzles, and reflective exercises.

In this way, the IVC module attempts to develop intellectual virtue by (a) introducing what virtue is (lecture videos defining and explaining the components of intellectual virtue); (b) explaining why it is important (videos and activities geared towards explaining how virtue is relevant to different disciplinary perspectives, as well as quiz questions testing virtue knowledge), and (c) detailing how to implement it in one’s own life (reflective exercises that ask individuals to think of areas of their life where they can implement virtue). Phrased another way, the mechanism by which the IVC is posited to operate involves the degree to which students acquire knowledge of intellectual virtue (what), grasp the significance of intellectual virtue (why), and understand what it takes to implement intellectual virtue (how). According to Besser’s theory outlined above, if these three criteria can be meaningfully and effectively targeted, individual growth in (intellectual) virtue is likely to follow suite.

In the pilot iteration of this intervention, Orona and Pritchard (2021) evaluated the IVC’s effect on two measures of intellectual curiosity and found positive preliminary

effects of 0.18 (Need for Cognition) and 0.13 (Epistemic Curiosity). To understand the mechanisms by which the IVC inculcates intellectual virtue, Orona (2021b) tested some of the theoretical links deemed essential for virtue learning (Besser, 2020). Consistent with Besser (2020)’s model and the broader intellectual character education stipulations, learning *what* intellectual virtue is and *why* it’s important was positively associated with increases in intellectual curiosity across a range of analytic approaches (Orona, 2021b).

Based upon this early work, the IVC has been expanded to include modules pertaining not only to intellectual curiosity, but also intellectual humility, integrity, and tenacity, as these are some of the commonly referenced intellectual virtues (Baehr, 2015; Pritchard, 2020). Therefore, in the current iteration of the IVC, students in the treatment group underwent videos, quizzes, brainstorming activities, and reflective exercises pertaining to each of these intellectual virtues.

Present study

The present study scales up the IVC evaluation, being among the first randomized controlled trials of an intentionally designed thinking disposition intervention aimed at improving university students intellectual virtue. Guided by the epistemic virtue frameworks indicating the significance of instruction on intellectual virtue development, Besser’s theory highlighting the what, why, and how of virtue learning, and the STRIVE-4 model (Fowers et al., 2021) hypotheses: “#23. Virtue acquisition can be fostered by well-designed, structured interventions” and “#24. Virtue-related behaviors can be increased with simple, short-term interventions”, we test the effect of the thinking disposition intervention on intellectual character development via a theoretically stipulated mechanism. Specifically, we specify latent variable mediation models using Besser’s learning components as mediator(s) transferring the effect of the treatment to growth on four key virtues: curiosity, tenacity, integrity, and humility. Thus, we test to research questions:

- RQ1: Does participation in the IVC increase intellectual virtue?
- RQ2: Do the components of virtue learning (e.g., understanding what virtue is, why it is important, and how to implement it) mediate the relationship between participation in the IVC and increases in intellectual virtue?

Methods

Participants

This study took place at a large public southern California research university. Instructors were contacted via email; those willing to embed the intervention in their course also agreed to offer extra credit to students for participating in the study. To be eligible for the study, participants needed to be: (a) enrolled in a participating course and (b) not previously exposed to the intervention. Initially, 806 undergraduates consented to participate. Due to attrition (i.e., either leaving the course or the study: 32%), eligibility criteria (of

those who completed the posttest, 23% were ineligible to be included in the study) and missing data (among those eligible, 15%), 424 students (treatment = 216; control = 208) had analyzable data. However, not all of these students had full administrative data to describe the sample. Those with complete data on demographics and academic records were $n = 361$ (treatment = 186; control = 175). We therefore specify latent variable models on the full sample (without covariate information; $n = 424$) and the analytic sample ($n = 361$) to compare how the drop in cases and inclusion of covariates impacts the results.

In Table 1, we present descriptive statistics by condition for the analytic sample, as well as mean differences. Across both conditions, most students were enrolled full-time during the term they participated (87%) and were female (> 60%). A little over 50% of the sample were first-generation college students, and about a one-third underrepresented minority students (URM: Hispanic, Black, and Native American). All imbalances between groups and pretest measures were tested with either a chi-square test (categorical) or a t -test (ordinal/numerical). No significant differences were found between the treatment and the control across any of the demographic, academic, and pretest survey measures.

Table 1 Descriptive Statistics by Treatment Condition

	Treatment M (SD)	Control M (SD)	Mean Difference
Demographics			
Full Load	0.87 (0.34)	0.86 (0.35)	0.01
Low-Income	0.38 (0.49)	0.33 (0.47)	0.05
First-Generation	0.52 (0.50)	0.54 (0.50)	-0.02
Underrepresented Minority (URM)	0.27 (0.44)	0.35 (0.48)	-0.08
Female	0.65 (0.48)	0.60 (0.49)	0.05
Academic Variables			
High School GPA	3.78 (0.85)	3.69 (1.04)	0.1
Entry Units	12.16 (2.87)	12.28 (3.04)	-0.12
Course Cluster: Arts & Humanities	0.57 (0.50)	0.55 (0.50)	0.02
Course Cluster: Education	0.13 (0.34)	0.14 (0.35)	-0.01
Course Cluster: STEM	0.30 (0.46)	0.30 (0.46)	0
Major Area: Arts & Humanities & Other	0.17 (0.37)	0.17 (0.37)	0
Major Area: Health Science	0.08 (0.26)	0.07 (0.25)	0.01
Major Area: Social Science	0.32 (0.47)	0.37 (0.48)	-0.05
Major Area: STEM	0.44 (0.50)	0.39 (0.49)	0.05
Term: Spring 2021	0.23 (0.42)	0.24 (0.43)	-0.01
Term: Fall 2021	0.40 (0.49)	0.35 (0.48)	0.05
Term: Winter 2022	0.22 (0.41)	0.23 (0.42)	-0.01
Term: Spring 2022	0.16 (0.36)	0.18 (0.38)	-0.02
Pretest Survey Measures			
Cognitive Reflection Test (CRT1)	1.54 (1.20)	1.62 (1.23)	-0.07
Process Data			
logMin	3.17 (1.66)	3.11 (1.42)	0.06
logMin2	2.94 (1.30)	2.98 (1.53)	-0.04
N	186	175	11

Note. The 1's and 2's after pretest scores refer to when the survey measure was assessed (1 = pretest; 2 = posttest). STEM = Science, Technology, Engineering and Mathematics; Full load: whether the student was taking a full load of credits during the term, Entry Units = entry credits completed up to the current term; log(min) = the log(minutes) it took students to complete both the pretest and posttest survey.

Procedure

We employ a pretest/posttest randomized controlled trial. Once students signed a consent form and completed a pre-survey, they were randomly assigned with 50% probability to either the thinking disposition intervention or a control condition. The control condition consisted of additional educational materials and exercises relating to the broad course domain for which the student was enrolled. For instance, students who signed-up for the extra-credit opportunity in their science course and who were assigned to the control condition received a science-focused module touching on a variety of fields to describe emergent phenomena. Students who signed-up for the extra-credit opportunity in their critical reasoning course and who were assigned to the control condition received additional materials developed by their instructor to further knowledge in the course topics. It's important to note that students were randomized *within* courses; this design feature obviates the plausibility of instructor and course topic effects (Orona, 2021a). In order to receive the post-survey, participants were required to send in proof of completion of the assigned modules via a snippet through email. All four intellectual virtues were assessed at both time points, while perceived virtue learning items were only assessed posttest. As these latter set of items ask explicit questions about intellectual virtues, this design feature was implemented to not reveal condition status to participants.

Measures

Intellectual virtue measures

To measure intellectual curiosity, we relied on the 18-item Need for Cognition (NFC) scale (Cacioppo & Petty, 1982). The measure has been shown to constitute a core aspect of epistemic curiosity (Powell et al., 2016). To measure intellectual humility, we deployed the 6-item scale developed by Leary et al. (2017). The IH scale has been shown to be highly correlated with other intellectual virtues and exhibit non-overlapping associations with personality (Leary et al., 2017). To measure intellectual integrity, the researchers devised their own 6-item scale, given no adequate preexisting scale. The items were devised to measure one's willingness to be intellectually honest, despite personal gain. Intellectual tenacity was also developed by the researchers, as no adequate preexisting scale was found. The construct is understood as applying effort towards intellectual goals despite the presence of obstacles. All scale response options were measured on a 5-point scale ranging from 1 = "Extremely Uncharacteristic" to 5 = "Extremely Characteristic".

Nine items were developed to measure Besser's virtue learning (BVL) components; specifically, items were designed to track students' subjective valuation of what intellectual virtue is, why it's important, and how to exercise it. Three items were used per learning type (e.g., what, why, how), each of which were positioned on a 5-point scale ranging from 1 = "Disagree Strongly" to 5 = "Strongly Agree." Full items for all scales can be found in the appendix.

Control variables

Control variables include demographic, academic, and additional pretest scores, which can be viewed in Table 1. Demographic and academic variables were obtained from university administrative records. The additional pretest measures include the 3-item cognitive reflection test (Frederick, 2005, p. CRT1: Cronbach's $\alpha = 0.75$) and the (log) minutes students spent on the pretest and posttest obtained from the survey software. The reason for the inclusion of the time spent on the surveys is because low-stakes testing/surveying studies are prone to elicit hasty and sometimes thoughtless responses from participants (Liu et al., 2012). Recent work shows that response times can provide valuable information on survey quality (Lundgren & Eklöf, 2023). Thus, by controlling total survey time, we are holding constant a proxy for the quality of responses from a student.

Data analysis

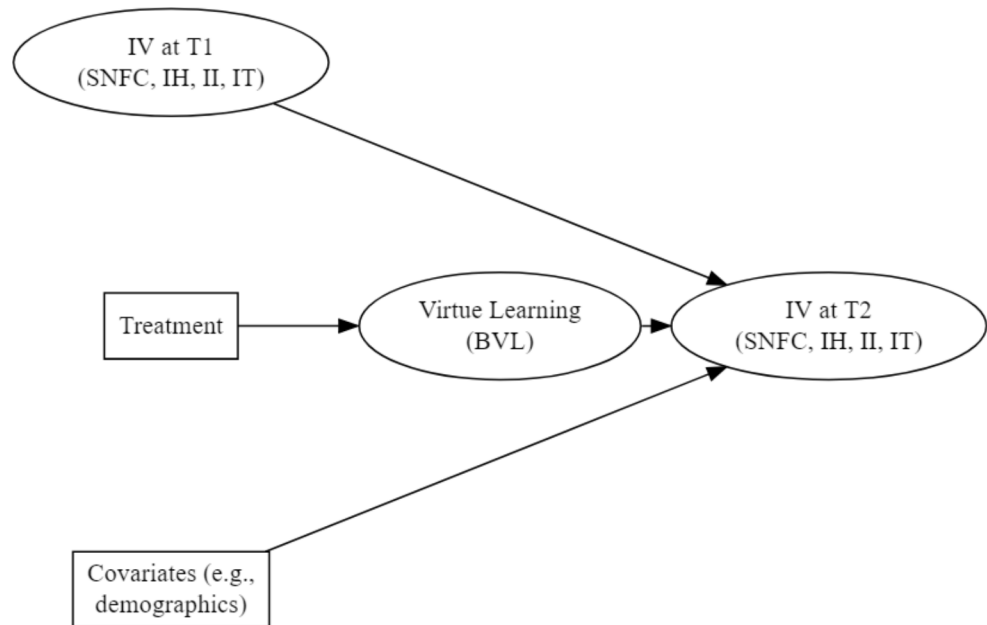
Prior to answering our main research questions, we test the adequacy of the intellectual virtue scales and the self-reported learning measures by conducting a series of measurement models. First, we begin with an exploratory factor analysis (EFA) of all the items for the intellectual virtue scales at time point 1 (36 items). Using criteria considering low loadings (< 0.40) and cross-loadings (> 0.30) outlined by Howard (2016), we drop items from the EFA. Then, we perform confirmatory factor analysis (CFA) to test configural, weak, and strong measurement invariance across the treatment groups at time point (1) We do the same invariance testing by group at time point (2) Finally, we do the same invariance testing across time points.

Since the Besser virtue learning measures were only assessed at time point 2, and because we have a theory-inspired latent variable structure, we begin with a CFA model. We then test a series of models to assess model fit. We also test configural, weak, and strong measurement invariance across the treatment groups.

For our main research question(s), we use two different strategies. First, we test the direct effects of the intervention on the IV variables using simple linear regression, reporting the coefficients and F -tests associated with each model. We use change score models for the virtue variables for interpretability. (Latent variable models testing direct effects are shown in the appendix.) Second, to test mediation, we build upon the invariance models in the preliminary analysis and specify cross-lagged, structural equation models (SEM) holding constant the time 1 intellectual virtue scores and maintaining the same invariance constraints of the strictest model obtained. We use SEM models to specify a mediation model with Besser's virtue-learning measures as the mediator between the treatment and the intellectual virtues. This type of model allows us to estimate the indirect effect of the treatment on the intellectual virtues via the intervening mechanisms of virtue learning. Figure 1 shows the general form of the mediation model.

Additionally, we specify models with and without covariate variables. The model with covariates includes a host of pretest characteristics (see Measures section above), a strategy proposed by Gelman et al. (2020) to stabilize estimates. Thus, while we report models with and without covariates, we focus on the former. We also report standardized coefficients of latent variables (mean of 0 and standard deviation of 1); thus, regression weights from the treatment indicator to the outcomes can therefore be interpreted akin to Cohen's d .

Fig. 1 Cross-lagged panel model with latent variable mediation. IV = Intellectual virtue scores; T1 = time-point 1/pretest; T2 = time-point 2/posttest; BVL = Besser’s Learning Index; SNFC = short Need for Cognition; IH = Intellectual Humility; II = Intellectual Integrity; IT = Intellectual Tenacity



Results

Preliminary analysis

The Kaiser-Meyer-Olkin (KMO) measure indicated good sampling adequacy, $KMO=0.87$. All individual items had KMO values above the 0.50 acceptable limit proposed by Hair et al. (2006). The Bartlett’s Test of Sphericity was $\chi^2 = 4854.76$ ($df = 630$), $p = 0$, indicating the data are suitable for reduction. Finally, the determination of the correlation matrix was positive, indicating the ability to extract common variance. The scree plot indicated a four-factor solution with the full 36 items included. After specifying a four-factor model and pruning items that had low loading and high cross-loading, we re-ran the EFA with varimax rotation. The four factors were a shortened NFC (SNFC; 8 items), IH (6 items), IT (3 items), and II (3 items). For time point 1 (T1) and time point 2 (T2), each scale exhibited adequate to strong reliability per Cronbach’s α : SFNC (T1 = 0.81; T2 = 0.83); IH (T1 = 0.84; T2 = 0.88); IT (T1 = 0.83; T2 = 0.82); II (T1 = 0.75; T2 = 0.71). Table 2 presents the factor loadings from the final solution (scree plot presented in the appendix).

Table 3 presents the results of the invariance testing. Both configural (e.g., model fit across groups), weak (e.g., factor loadings equal across groups), and strong (e.g., item intercepts equal across groups) invariance were met between the treatment and control groups for both time 1 and time 2, as indicated by the non-significant p -value(s). We also tested the same set of invariances across time points (before and after the intervention). Across time points, configural invariance was met but weak was not achieved. A

typical approach when invariance is not met is to seek partial invariance (e.g., allowing some item loadings/intercepts to be freely estimated across groups/time points). Thus, after identifying the item causing a decrease in fit (one item from the NFC scale) and allowing it to be freely estimated across time points, partial weak invariance was achieved. Similarly, we found one problematic item intercept for the test of strong invariance (another item from the NFC scale) that, once allowed to be freely estimated across time points, partial weak invariance was achieved.

For the virtue learning measures, we compared a variety of models, including a model specifying a meta-factor for virtue learning composed of three first-level factors corresponding to the three aspects of virtue learning: what, why, and how. This model, which included model implied correlated uniqueness, had the best global fit (CFI=0.985; TLI=0.972; SRMR=0.026; RMSEA=0.071) and the lowest relative fit (e.g., using BIC and AIC metrics) of the models compared (competing models are shown in the supplemental material). This model and the corresponding factor loadings are shown in Fig. 2. Thus, group invariance was tested on this model, as presented in Table 4. Both configural (e.g., model fit across groups) and weak (e.g., factor loadings equal across groups) invariance were met, as indicated by the non-significant p -value. However, one item’s intercept required to be freely estimated across treatment status; thus, partial strong invariance was achieved.

Table 2 Factor Loadings for Intellectual Virtue Scales

	SNFC	IH	II	IT
	$\alpha =$	$\alpha =$	$\alpha =$	$\alpha =$
	0.81	0.81	0.75	0.83
nfc1_2: I like to have the responsibility of handling a situation that requires a lot of thinking.	0.46	0.17	0.13	0.27
nfc1_3: Thinking is not my idea of fun.	0.64	0.16	-0.03	0.09
nfc1_4: I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.	0.78	0.05	0.07	0.13
nfc1_5: I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.	0.69	0.08	0.06	0.15
nfc1_7: I only think as hard as I have to.	0.45	-0.06	0.15	0.11
nfc1_9: I like tasks that require little thought once I've learned them.	0.54	-0.1	0.06	0.05
nfc1_12: Learning new ways to think doesn't excite me very much.	0.5	0.11	0.04	0.17
nfc1_17: It's enough for me that something gets the job done; I don't care how or why it works.	0.52	0.05	0.05	0.16
ih1_1: I question my own opinions, positions, and viewpoints because they could be wrong.	-0.05	0.57	0	0.03
ih1_2: I reconsider my opinions when presented with new evidence.	0.08	0.66	0.1	0.12
ih1_3: I recognize the value in opinions that are different from my own.	0.13	0.74	0.06	0.14
ih1_4: I accept that my beliefs and attitudes may be wrong.	0.02	0.77	0.09	0.03
ih1_5: In the face of conflicting evidence, I am open to changing my opinions.	0.03	0.74	0.09	0.1
ih1_6: I like finding out new information that differs from what I already think is true.	0.15	0.54	0.05	0.24
ii1_4: If I know that another student is being intellectually dishonest, I will encourage them to stop.	0.11	-0.01	0.52	0.14
ii1_5: If I notice someone intentionally deceiving someone, I will speak.	0.08	0.21	0.84	0.1
ii1_6: If I notice someone unintentionally deceiving someone, I will speak.	0.09	0.12	0.71	0.09
it1_1: When I encounter intellectually demanding tasks, I continue to push myself even though it is difficult.	0.26	0.22	0.11	0.62
it1_3: Even if it takes a long time and lots of effort, I will not give up my intellectual interest	0.22	0.17	0.08	0.73
it1_4: If it gets more difficult to pursue my intellectual interest	0.24	0.12	0.15	0.74
it1_6: I can motivate myself to continue pursuing my intellectual interest	0.17	0.08	0.12	0.62

Note. SNFC = shortened Need for Cognition; IH = Intellectual Humility; II = Intellectual Integrity; IT = Intellectual Tenacity

Table 3 Measurement Invariance of Intellectual Virtues (across groups and time)

Invariance by Treatment Status at Time 1							
	<i>df</i>	AIC	BIC	χ^2	χ^2 diff	<i>df</i> diff	<i>p</i>
Configural Invariance	366	21804.41	22363.27	593.27			
Weak Invariance	383	21785.29	22275.31	608.15	12.03	17	0.798
Strong Invariance	400	21778.41	22199.58	635.27	27.04	17	0.057
Invariance by Treatment Status at Time 2							
	<i>df</i>	AIC	BIC	χ^2	χ^2 diff	<i>df</i> diff	<i>p</i>
Configural Invariance	366	21013.04	21571.91	585.95			
Weak Invariance	383	21009.01	21499.02	615.91	24.15	17	0.115
Strong Invariance	400	20996.96	21418.13	637.87	22.53	17	0.165
Invariance Across Time Points							
	<i>df</i>	AIC	BIC	χ^2	χ^2 diff	<i>df</i> diff	<i>p</i>
Configural Invariance	366	42900.37	43555.21	751.85			
(Partial) Weak Invariance	382	42895.34	43474.26	778.82	22.01	16	0.143
(Partial) Strong Invariance	398	42878.76	43381.76	794.24	15.46	16	0.491

Note. *df*=degrees of freedom; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Fig. 2 CFA model for the Besser Virtue Learning index. A meta-factor model with first-order factors (what, why, how) are comprised of nine manifest variables

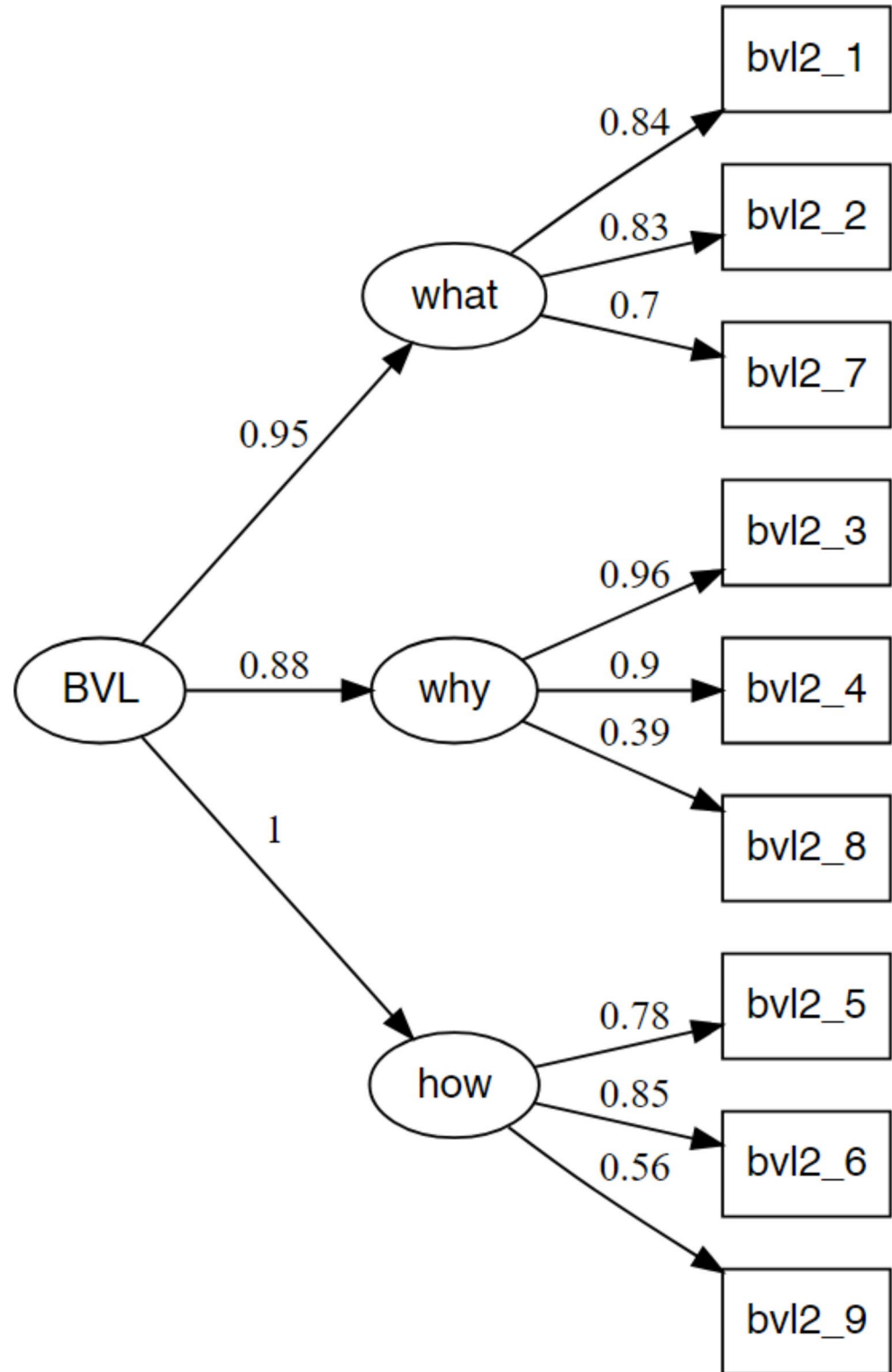


Table 4 Besser Virtue Learning Invariance Testing Models Group Comparison

	<i>df</i>	AIC	BIC	χ^2	χ^2 diff	<i>df</i> diff	<i>p</i>
Configural Invariance	32	7934.62	8240.59	74.203			
Weak Invariance	40	7934.3	8208.06	89.8791	13.7796	8	0.08769
(Partial) Strong Invariance	44	7933.12	8190.77	96.6977	5.21768	4	0.26568

Note. *df*=degrees of freedom; AIC= Akaike Information Criterion; BIC= Bayesian Information Criterion.

RQ1: does participation in the IVC increase intellectual virtue?

Figure 3 shows the mean scores for each IV variable by time point and condition. Descriptively, we see that the differences between conditions appear minimal for all the virtues.

Table 5 presents the formal tests of these comparisons, showing direct effects of the treatment on the standardized IV change scores. The treatment has a medium-sized effect on both SNFC (0.20, $p < .001$) and IT (0.21, $p < .05$). Interestingly, all effects were centered on positive values except the II change score, though this wasn't significant ($p > .05$).

Relatedly, the F -test is significant for all models except IH and II.

RQ2: does virtue learning mediate the relationship between IVC participation and virtue development?

Table 6 presents the results of the main analysis. First, we see that the treatment had a very large direct effect on the BVL meta-factor in the model without covariates ($n = 424$) conducted on the full sample ($\beta = 0.72, p < .001$) and the model with covariates ($n = 361$) conducted on the sub-sample ($\beta = 0.77, p < .001$).

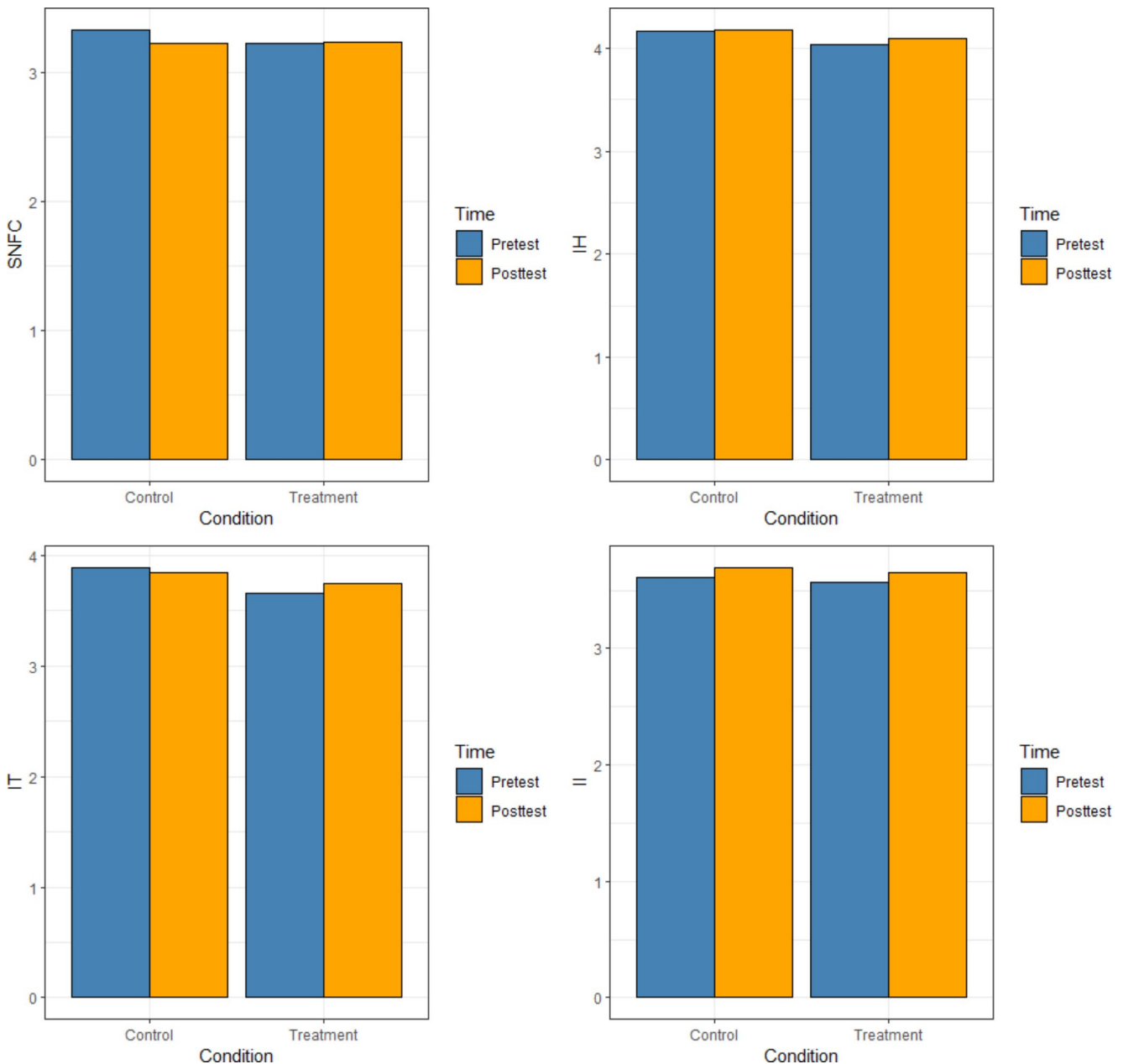


Fig. 3 Mean plots for each intellectual virtue before and after treatment by condition

Table 5 Manifest Variable Models: Direct Effects of the Treatment

	SNFC Change Score	IT Change Score	IH Change Score	II Change Score
Treatment: IVC Condition (95% CI Lower -Upper)	0.202* (0.013–0.392)	0.211* (0.022–0.401)	0.068 (-0.123–0.258)	-0.016 (-0.206–0.175)
Intercept	-0.103	-0.108	-0.034	0.008
Observations	424	424	424	424
R ²	0.010	0.011	0.001	0.0001
F Statistic (df= 1; 422)	4.376*	4.782*	0.485	0.026

Note: * $p < .05$; ** $p < .01$; *** $p < .001$. SNFC = shortened Need for Cognition; IH = Intellectual Humility; II = Intellectual Integrity; IT = Intellectual Tenacity; df = degrees of freedom.

Table 6 Latent Variable Mediation Models

Parameter	Model 1 ($n = 424$)			Model 2 ($n = 361$)		
	Estimate ^a (SE)	z -value	CI (Lower – Upper)	Estimate ^a (SE)	z -value	CI
Direct Effect on Mediator (BVL)	0.723*** (0.093)	6.96	(0.541–0.905)	0.770*** (0.099)	6.96	
Indirect Effect on SNFC	0.134** (0.045)	2.88	(0.045–0.222)	0.111* (0.051)	2.18	
Indirect Effect on IH	0.161** (0.052)	2.91	(0.059–0.263)	0.145** (0.055)	2.49	
Indirect Effect on II	0.218*** (0.061)	3.44	(0.097–0.338)	0.189** (0.063)	2.89	
Indirect Effect on IT	0.239*** (0.057)	4.4	(0.127–0.35)	0.259*** (0.064)	4.11	

Note. ^a = estimates are based upon standardized latent variables. Thus, the direct effect of the treatment on the mediator (BVL = Besser's Virtue Learning) can be interpreted as an effect size. SE = Standard Error. Model 1 does not have any covariates besides T1 virtue scores and is conducted on the full sample. The fit for model 1 is as follows: $\chi^2 = 1999.015$ (df = 1,232, $p < .05$); CFI = 0.937; TLI = 0.932; SRMR = 0.082; RMSEA = 0.0035; AIC = 48705.86; BIC = 49494.93. Model 2 includes covariates (e.g., demographic, academic, and CRT and minutes on test). The fit for model 2 is as follows: $\chi^2 = 2825.319$ (df = 2,031, $p < .05$); CFI = 0.956; TLI = 0.953; SRMR = 0.075; RMSEA = 0.0031; AIC = 41035.41; BIC = 42055.4. BVL = Besser's Virtue Learning index; SNFC = shortened Need for Cognition; IH = Intellectual Humility; II = Intellectual Integrity; IT = Intellectual Tenacity; df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation; BIC = Bayesian Information Criterion; CI = Confidence Interval.

Additionally, all indirect effects from the treatment to the four intellectual virtues via BVL were statistically significant, $p < .05$. Furthermore, the estimates were stable across the models/samples presented, with minimal difference in the size or significance of the estimates. The intellectual virtue most impacted by the treatment through BVL was IT, irrespective of the model without ($\beta = 0.24$, $p < .001$) and with covariates ($\beta = 0.26$, $p < .001$). The intellectual virtue least impacted by the treatment through BVL was SNFC, irrespective of the model without ($\beta = 0.13$, $p < .01$) and with covariates ($\beta = 0.11$, $p < .05$).

To visualize the extent to which change in the virtue scores correlates with virtue learning, Fig. 4 presents the bivariate correlations between the factor change scores and the extracted factor scores for the BVL latent variable. All BVL associations with virtue change scores were statistically significant ($p < .001$), except for intellectual humility

(IH), which approximated significance at the conventional threshold ($p = .0502$).

Discussion

In line with the budding interest in both virtue science and intellectual character education in higher education, the purpose of the present study was to evaluate the impact of a thinking disposition intervention on intellectual virtue development and learning. Using a randomized controlled trial design, we measured and assessed the direct and indirect effect of the IVC on four key virtues: curiosity, humility, integrity, and tenacity through subjective virtue learning measures. The experiment generated generally positive results for the intervention—though there are some notable aspects suggesting limited effectiveness.

For example, the change score regression models revealed that there was no significant impact on II or IH. The intervention did, however, have large impacts on virtue learning, with effect sizes approaching a full standard deviation unit. Moreover, we observed statistically significant indirect effects of participation for every one of the four virtues across the full sample and the sub-sample with covariate variables included. Given these findings, the results of this study have implications for virtue science and the design of intellectual character interventions.

Virtue science

The study results have clear implications for models of (epistemic) virtue. First, the intervention was effective at promoting knowledge of what intellectual virtue is, why it is important, and how to implement it, suggesting the importance of instruction in virtue learning. Second, based on the

current study’s results, Besser’s theory of virtue learning was well supported, showing that learning the *what, why, and how* of virtue leads to greater growth in intellectual virtue. Fourth, guided by the STRIVE-4 framework, this study communicates with criteria for virtue science. This contribution can be realized by the study results that show virtue acquisition can be fostered by well-designed, structured interventions. Furthermore, given an academic term at the study university entails a total of 10 weeks, and students complete the intervention in approximately 7, then, depending on one’s frame of reference, this may qualify as a short-term intervention. These two points relate to the two STRIVE-4 hypotheses regarding interventions and provide a basis to begin unpacking the specific components that foster virtue development.

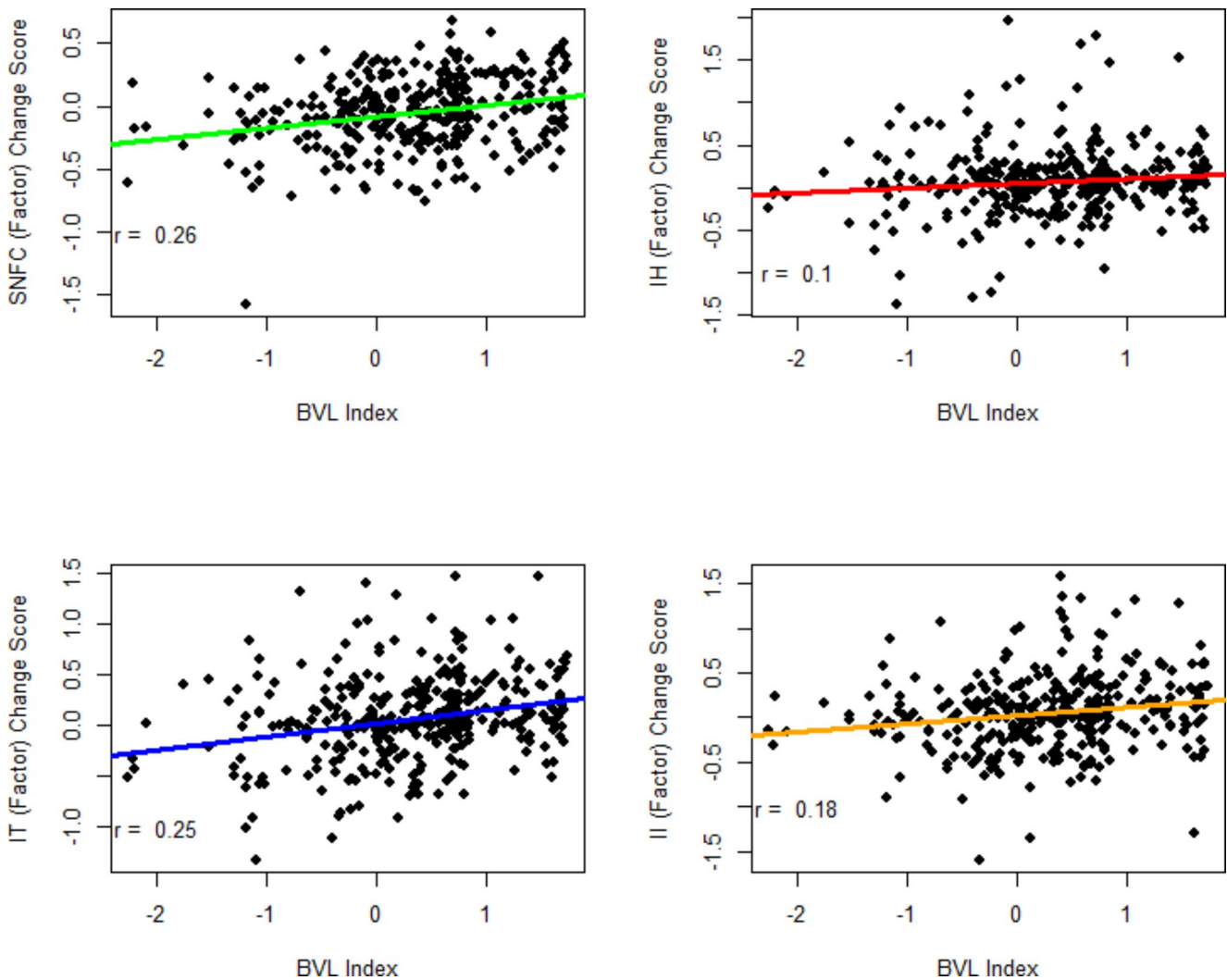


Fig. 4 Bivariate relations between intellectual virtue change scores (computed from extracted factor scores) and the virtue learning (Besser Virtue Learning = BVL) variable

Design and evaluation of character interventions

Ruch et al. (2020) notes that most character interventions emphasize outcomes *other* than character, noting the need for studies showing their efficacy for character development. Insofar as the current study measures are adequate representations of intellectual character, our study contributes to the knowledge base for the efficacy of character education interventions on character attributes. Moreover, it's in line with recent suggestions regarding rigor in evaluating character education programs (McGrath, 2022). We also found that the treatment effects on the (shortened) Need for Cognition scale (0.20) and the intellectual tenacity (0.21) scale corresponded well with the effect sizes of previous character development programs (0.24, Brown et al., 2022), and the pilot iteration of the thinking disposition intervention (0.18, Orona & Pritchard, 2021).

Thus, another implication is that not all character traits can be impacted equally, despite equal intervention. For example, we found that intellectual curiosity—as measured by NFC—and tenacity were more impacted than humility and integrity, yet the treatment emphasized each trait equally. There are several possible explanations for this. First, it may be that the IVC modules relating to the former two traits are somehow better designed, thus resulting in stronger gains in these dispositions. Second, it may be that the virtues themselves are sufficiently different, such that the curiosity and tenacity are more fluid and easily targeted than integrity and humility, which may be more deeply ingrained character attributes and thus require more intense intervention. Third, there may be a ceiling effect: participants may have been already too high on the former traits to grow. This may be the case for intellectual humility, which was the only virtue that had a strong negative skew in its distribution (histograms of every outcome variable can be found in the supplementary material).

And most importantly, there is the possibility of measurement error. However, insofar as reliability estimates are concerned, no clear pattern was observed in the traits that were and were not impacted. For example, the two significantly impacted virtues included the most established scale administered (e.g., NFC) and a researcher-developed scale (tenacity)—each with high to adequate Cronbach α values. Likewise, the two virtues not statistically significantly impacted included one previous existing scale (humility), and one researcher-developed scale (integrity)—exhibiting high to less than adequate Cronbach α values, respectively. Still, only the NFC has undergone extensive psychometric testing and validation, with over a 40-year research base (e.g., Cacioppo & Petty, 1982; Lavrijsen et al., 2021). We further expound on measurement issues as they relate to limitations and future directions in the section below.

Limitations and future directions

The limitations largely concern measurement, external validity, and follow-up. Ideally, measurements beyond self-report are desired¹ (Maul, 2017). But insofar as psychometrics is concerned, no clear pattern was observed in the traits that were and were not directly impacted. For example, the two significantly impacted virtues included the most established scale administered (e.g., NFC) and a researcher-developed scale (tenacity)—each with high to adequate Cronbach α values. Still, advancing situated, contextual measures with enhanced predictive validity will require extensive conceptual and empirical research moving well beyond self-report scales (Ng & Tay, 2020). Another limitation is external validity, which could be enhanced with a multi-campus experiment examining the effects of IVC within and across diverse institutions (and populations). Finally, a follow-up study would greatly strengthen our understanding of the lasting effect of the IVC intervention. Together, such data would be vital in gaining knowledge of if and when supports and boosters may be relevant to sustain fundamental intellectual character development.

Conclusion

Intellectual virtue development *appears* to be a feasible aim when intervention design is guided by philosophical, psychological, and pedagogical perspectives. This study demonstrated the effect of intellectual virtue education on learning about virtue, which in turn was related to virtue development. Given wide-spread appeal in developing life-long learners with reflective thinking habits and good epistemic hygiene, researchers and funders may consider investing time and resources in intellectual character education as a potential avenue towards a society of virtuous reasoners.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12144-023-05005-1>.

Acknowledgements This work was supported by the Educational Research Initiative under IRB Grant #486 Evaluation of the Anteatr Virtues Curriculum Project.

¹ While not motivated by the current framework, we provide an exploratory test of whether the IVC improves final course grades. The intervention did not have a significant impact, $d=0.008$, $p>.05$. This is unsurprising, as students in both conditions received extra credit, and the control group received additional, course-relevant content as part of the study than the treatment. We also tested overall satisfaction with the extra credit opportunities (for both treatment and control groups); the IVC participants were significantly more satisfied with their experience (meaning, the IVC participants enjoyed the IVC more than the control group enjoyed the control condition), $d=0.24$, $p<.05$.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The authors report there are no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anjum, A., & Amjad, N. (2021). Values in action inventory of strengths: Development and validation of short form-72 in Urdu. *Current Psychology*, *40*, 2039–2051.
- 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.
- 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*. 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 Slote, & E. Sosa (Eds.), *Moral and intellectual virtues in western and chinese philosophy* (pp. 86–98). Routledge. Chienkuo Mi Michael.
- Baehr, J. (2021). *Deep in thought: A practical guide to teaching for intellectual virtues*. ERIC.
- Besser, L. L. (2020). Learning virtue. *Journal of Moral Education*, *49*(3), 282–294.
- Brown, M., McGrath, R. E., Bier, M. C., Johnson, K., & Berkowitz, M. W. (2022). A comprehensive meta-analysis of character education programs. *Journal of Moral Education*, 1–20.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.
- Fischer, C., McPartlan, P., Orona, G. A., Yu, R., Xu, D., & Warschauer, M. (2022). Salient syllabi: Examining design characteristics of science online courses in higher education. *PLoS One*, *17*(11), e0276839.
- Fowers, B. J., Carroll, J. S., Leonhardt, N. D., & Cokelet, B. (2021). The emerging science of virtue. *Perspectives on Psychological Science*, *16*(1), 118–147.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.
- Greco, J. (2000). Two kinds of intellectual virtue. *Philosophy and Phenomenological Research*, *60*(1), 179–184.
- Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 45–53.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* 6th edition. Pearson Prentice Hall. New Jersey. *Humans: Critique and Reformulation. Journal of Abnormal Psychology*, *87*, 49–74.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127.
- Hong, H. Y., & Lin-Siegler, X. (2012). How learning about scientists' struggles influences students' interest and learning in physics. *Journal of Educational Psychology*, *104*(2), 469.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62.
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412.
- 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.
- Kotsonis, A. (2020). What can we learn from plato about intellectual character education? *Educational Philosophy and Theory*, *52*(3), 251–260.
- Kotzee, B., Carter, J. A., & Siegel, H. (2019). Educating for intellectual virtue: A critique from action guidance. *Episteme*, 1–23.
- Krumrei-Mancuso, E. J., Haggard, M. C., LaBouff, J. P., & Rowatt, W. C. (2020). Links between intellectual humility and acquiring knowledge. *The Journal of Positive Psychology*, *15*(2), 155–170.
- 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.
- 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.
- Lian, L., & You, X. (2017). Specific virtues as predictors of smartphone addiction among chinese undergraduates. *Current Psychology*, *36*(2).
- Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, *41*(9), 352–362.
- Lundgren, E., & Eklöf, H. (2023). Questionnaire-taking motivation: Using response times to assess motivation to optimize on the PISA 2018 student questionnaire. *International Journal of Testing*, 1–26.
- Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary research and perspectives*, *15*(2), 51–69.
- McGrath, R. E. (2022). Some key issues in the evaluation of character education. *Journal of Education*, *202*(2), 181–184.
- 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.
- 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.

- Ng, V., & Tay, L. (2020). Lost in translation: The construct representation of character virtues. *Perspectives on Psychological Science*, *15*, 309–326. <https://doi.org/10.1177/1745691619886014>.
- 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., & Pritchard, D. (2021). Inculcating curiosity: Pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment & Evaluation in Higher Education*, 1–15.
- Orona, G. A., Eccles, J. S., Zitzmann, S., Fischer, C., & Arum, R. (2023). Cognitive development in undergraduate emerging adults: How course-taking breadth supports skill formation. *Contemporary Educational Psychology*, 102206.
- 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.
- Porter, T., Schumann, K., Selmezy, D., & Trzesniewski, K. (2020). Intellectual humility predicts mastery behaviors when learning. *Learning and Individual Differences*, *80*, 101888.
- Powell, C., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences*, *95*, 147–151.
- Pritchard, D. (2020). Educating for intellectual humility and conviction. *Journal of Philosophy of Education*, *54*(2), 398–409.
- Quinlan, K. M. (2019). What triggers students' interest during higher education lectures? Personal and situational variables associated with situational interest. *Studies in Higher Education*, *44*(10), 1781–1792.
- Ruch, W., Niemiec, R. M., McGrath, R. E., Gander, F., & Proyer, R. T. (2020). Character strengths-based interventions: Open questions and ideas for future research. *The Journal of Positive Psychology*, *15*(5), 680–684.
- Rudolph, J., Greiff, S., Strobel, A., & Preckel, F. (2018). Understanding the link between need for cognition and complex problem solving. *Contemporary Educational Psychology*, *55*, 53–62.
- Schwartz, B. (2020). Science, scholarship, and intellectual virtues: A guide to what higher education should be like. *Journal of Moral Education*, 1–12.
- Sosa, E. (1980). The raft and the pyramid: Coherence versus foundations in the theory of knowledge. *Midwest studies in philosophy*, *5*(1), 3–26.
- Sosa, E. (1985). Knowledge and intellectual virtue. *The Monist*, *68*(2), 226–245.
- Šrol, J. (2018). These problems sound familiar to me: Previous exposure, cognitive reflection test, and the moderating role of analytic thinking. *Studia Psychologica*, *60*(3), 195–208.
- Villacis, J. L., Naval, C., & De la Fuente, J. (2022). Character strengths, moral motivation and vocational identity in adolescents and young adults: A scoping review. *Current Psychology*, 1–16.
- Vossoughi, S., Davis, N. R., Jackson, A., Echevarria, R., Muñoz, A., & Escudé, M. (2021). Beyond the binary of adult versus child centered learning: Pedagogies of joint activity in the context of making. *Cognition and Instruction*, *39*(3), 211–241.
- Yuretich, R. F., Khan, S. A., Leckie, R. M., & Clement, J. J. (2001). Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course. *Journal of Geoscience Education*, *49*(2), 111–119.
- Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.
- Zhang, K., Shi, J., Wang, F., & Ferrari, M. (2022). Wisdom: Meaning, structure, types, arguments, and future concerns. *Current Psychology*, 1–22.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.