Intellectual Humility, Arrogance, and Openness: Investigating the Psychometric Properties of a Self-Report Measure of Intellectual Virtue

A thesis submitted by

Elise D. Murray

In partial fulfillment of the requirements for the degree of

Master of Arts

in

Child Study and Human Development

Tufts University

May 2017

ADVISOR: Richard M. Lerner

Abstract

Intellectual virtues are important characteristics for pursuing life-long learning. Unfortunately, some intellectual virtues are not conceptualized or empirically assessed using a life-long, or developmental, framework; one such intellectual virtue is that of intellectual humility (IH). The purpose of this study is to begin to contribute to the development of a change-sensitive, self-report measure of IH. Using a sample of first-year United States Military Academy cadets (N =1,257), the study assessed the factor structure of a self-report adaptation of an already existing other-report IH measure, the Intellectual Humility Scale (IHS; McElroy et al., 2014). In addition, the present analyses assessed whether the measure was invariant by gender (i.e., men and women). The self-report IHS showed best model fit with a two-factor structure, and this factor structure was invariant by gender, with only a minor modification. By establishing the foundation for a psychometrically sound self-report measure of IH, researchers can begin to measure IH developmentally, to assess how this intellectual virtue may co-act within the individual⇔context process of life-long learning.

Acknowledgements

This research was supported by a grant from the Templeton Religion Trust. The author would like to thank the members of her thesis committee, Sara K. Johnson and Scott Seider, as well as her committee chair, Richard M. Lerner, for their feedback. She would also like to thank the Principal Investigator of Project Arete, Kristina S. Callina, for her guidance and mentorship throughout the process of devising and writing this thesis.

Table of Contents

Table 1

Intellectual Humility Scale Item Stems

Subscale	Number	Stem
Intellectual Arrogance		
	1	I often become angry when my ideas are not implemented.
	2	I value winning an argument over maintaining a relationship.
	3	I always have to have the last word in an argument.
	4	I get defensive if others do not agree with me.
	5	I become angry when my advice is not taken.
	6	I have little patience for others' beliefs.
	7	I act like a know-it-all.
	8	I often point out others' mistakes.
	9	I make fun of people with different viewpoints.
Intellectual Openness		
	10	I seek out alternative viewpoints.
	11	I encourage others to share their viewpoints.
	12	I enjoy diverse perspectives.
	13	I am open to competing ideas.
	14	I am good at mediating controversial topics.
	15	I am good at considering the limitations of my perspective.
	16	I am open to others' ideas.

Table 2

Item Level Descriptive Statistics

						Skewness		Kurtosis	
Item	*N*	Min	Max	Mean	*SD*	Statistic	*SE*	Statistic	*SE*
1*	1254	1.00	5.00	3.63	.81	-.351	.069	-.161	.14
2*	1254	1.00	5.00	4.17	.75	-.817	.069	.725	.14
3*	1255	1.00	5.00	3.69	.92	-.533	.069	-.218	.14
4*	1254	1.00	5.00	3.60	.90	-.375	.069	-.477	.14
5*	1253	1.00	5.00	3.75	.84	-.555	.069	.088	.14
6*	1255	1.00	5.00	4.34	.71	-1.043	.069	1.509	.14
7*	1255	1.00	5.00	4.11	.87	-.874	.069	.279	.14
8*	1251	1.00	5.00	3.54	.93	-.326	.069	-.608	.14
9*	1249	1.00	5.00	4.32	.77	-1.105	.069	1.056	.14
10	1255	1.00	5.00	3.82	.76	-.710	.069	1.027	.14
11	1252	1.00	5.00	4.01	.69	-.518	.069	.692	.14
12	1249	1.00	5.00	4.08	.70	-.520	.069	.508	.14
13	1250	1.00	5.00	4.08	.62	-.469	.069	1.208	.14
14	1248	1.00	5.00	3.69	.83	-.310	.069	-.187	.14
15	1246	1.00	5.00	3.78	.75	-.515	.069	.367	.14
16	1247	1.00	5.00	4.19	.61	-.556	.069	1.515	.14
IH**	1257	2.25	5.00	3.93	.41	-.066	.069	.166	.14

*These items were part of the Intellectual Arrogance subscale and were reverse coded for the one-factor models. The items were analyzed in their original form for the two-factor models.
**IH = Intellectual Humility (as indicated by the Intellectual Humility Scale, McElroy et al., 2014)

Table 3

*Statistically Significant Correlations among Intellectual Humility Scale Items***

	1*	2*	3*	4*	5*	6*	7*	8*	9*	10	11	12	13	14	15	16
1*																
2*	.295															
3*	.324	.414														
4*	.405	.248	.395													
5*	.482	.200	.304	.416												
6*	.265	.304	.239	.275	.320											
7*	.231	.231	.288	.284	.237	.256										
8*	.170	.189	.252	.255	.251	.199	.382									
9*	.213	.236	.234	.241	.183	.381	.302	.232								
10	.126	.087	.138	.133	.138	.253	.102		.177							
11	.125	.157	.128	.154	.084	.273	.114		.261	.421						
12	.145	.151	.174	.178	.151	.369	.136	.065	.294	.424	.546					
13	.168	.117	.171	.203	.162	.243	.081		.188	.341	.371	.485				
14			.097	.095		.176			.108	.206	.238	.249	.317			
15	.114	.149	.172	.132	.135	.172	.127	.076	.149	.216	.292	.258	.310	.364		
16	.199	.219	.233	.227	.212	.383	.216	.129	.295	.335	.402	.498	.438	.242	.318	
IH	.527	.488	.574	.582	.536	.602	.508	.437	.537	.473	.521	.580	.521	.380	.464	.608

*These items were part of the Intellectual Arrogance subscale and were reverse coded for the one-factor models. The items were analyzed in their original form for the two-factor models.

** $p<.05$

Table 4

Unstandardized Factor Loadings for Single-Factor Model with Uncorrelated Residuals

Item	Estimate	*p*	SE
1*	1.00	-	-
2*	0.88	0.00	.09
3*	1.23	0.00	.11
4*	1.22	0.00	.11
5*	1.05	0.00	.10
6*	1.26	0.00	.11
7*	1.01	0.00	.10
8*	0.77	0.00	.10
9*	1.13	0.00	.10
10	1.12	0.00	.11
11	1.20	0.00	.11
12	1.39	0.00	.12
13	1.11	0.00	.10
14	0.87	0.00	.10
15	0.97	0.00	.10
16	1.25	0.00	.10

*These items were part of the Intellectual Arrogance subscale and were reverse coded for the one-factor models. The items were analyzed in their original form for the two-factor models.

Table 5

Unstandardized Factor Loadings for Single-Factor Model with Correlated Residuals

Item	Estimate	*p*	SE
1*	1.00	-	-
2*	0.93	0.00	.11
3*	1.30	0.00	.14
4*	1.31	0.00	.13
5*	1.11	0.00	.10
6*	1.47	0.00	.14
7*	1.09	0.00	.13
8*	0.76	0.00	.12
9*	1.31	0.00	.14
10	1.37	0.00	.14
11	1.38	0.00	.14
12	1.66	0.00	.16
13	1.39	0.00	.14
14	1.12	0.00	.13
15	1.20	0.00	.13
16	1.53	0.00	.14

*These items were part of the Intellectual Arrogance subscale and were reverse coded for the one-factor models.
The items were analyzed in their original form for the two-factor models.

Table 6

Unstandardized Factor Loadings for Two-Factor Model with Uncorrelated Residuals

Item	Estimat	*p*	SE
Intellectual Arrogance			
1	1.00	-	-
2	0.80	0.00	.06
3	1.34	0.00	.08
4	1.16	0.00	.07
5	1.02	0.00	.07
6	0.85	0.00	.06
7	0.99	0.00	.07
8	0.83	0.00	.07
9	0.79	0.00	.06
Intellectual Openness			
10	1.00	-	-
11	1.06	0.00	.06
12	1.24	0.00	.07
13	0.98	0.00	.06
14	0.79	0.00	.07
15	0.80	0.00	.06
16	0.98	0.00	.06

Note: IA = Intellectual Arrogance; IO = Intellectual Openness

Table 7

Unstandardized Factor Loadings for Two-Factor Model with Correlated Residuals

Item	Estimate	*p*	SE
Intellectual Arrogance			
1	1.00	-	-
2	0.86	0.00	.08
3	1.22	0.00	.10
4	1.26	0.00	.09
5	1.08	0.00	.08
6	1.07	0.00	.09
7	1.06	0.00	.09
8	0.88	0.00	.09
9	0.99	0.00	.09
Intellectual Openness			
10	1.00	-	-
11	1.00	0.00	.06
12	1.20	0.00	.07
13	1.01	0.00	.06
14	0.83	0.00	.07
15	0.83	0.00	.07
16	1.03	0.00	.06

Table 8

Standardized Factor Loadings of Configural Two-Factor Model with Correlated Residuals by Gender

Item	Women			Men		
	Estimate	*p*	SE	Estimate	*p*	SE
Intellectual Arrogance						
1	.51	0.00	.04	.51	0.00	.03
2	.44	0.00	.04	.47	0.00	.03
3	.53	0.00	.04	.56	0.00	.03
4	.56	0.00	.04	.59	0.00	.03
5	.51	0.00	.04	.53	0.00	.03
6	.59	0.00	.04	.58	0.00	.03
7	.47	0.00	.04	.49	0.00	.03
8	.38	0.00	.04	.39	0.00	.03
9	.53	0.00	.04	.49	0.00	.03
Intellectual Openness						
10	.56	0.00	.03	.55	0.00	.03
11	.66	0.00	.04	.60	0.00	.02
12	.74	0.00	.03	.70	0.00	.02
13	.64	0.00	.03	.64	0.00	.02
14	.40	0.00	.04	.41	0.00	.03
15	.48	0.00	.04	.46	0.00	.03
16	.66	0.00	.03	.68	0.00	.02

Table 9

Model fit for between-group invariance tests by gender

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI	Pass? (ΔCFI ≤ .01)
Configural	604.45 (196)	<.001	0.058 (0.053 to 0.064)	0.911	0.890	
Loading	618.75 (210)	<.001	0.057 (0.051 to 0.062)	0.911	0.898	Pass (ΔCFI = .000)
Intercept	684.15 (224)	<.001	0.058 (0.053 to 0.063)	0.899	0.892	Fail (ΔCFI = .012)
a. Full						
b. Partial*	667.053 (223)	<.001	0.057 (0.052 to 0.062)	0.903	0.895	Pass (ΔCFI = .008)

*Note. Partial = Item 12

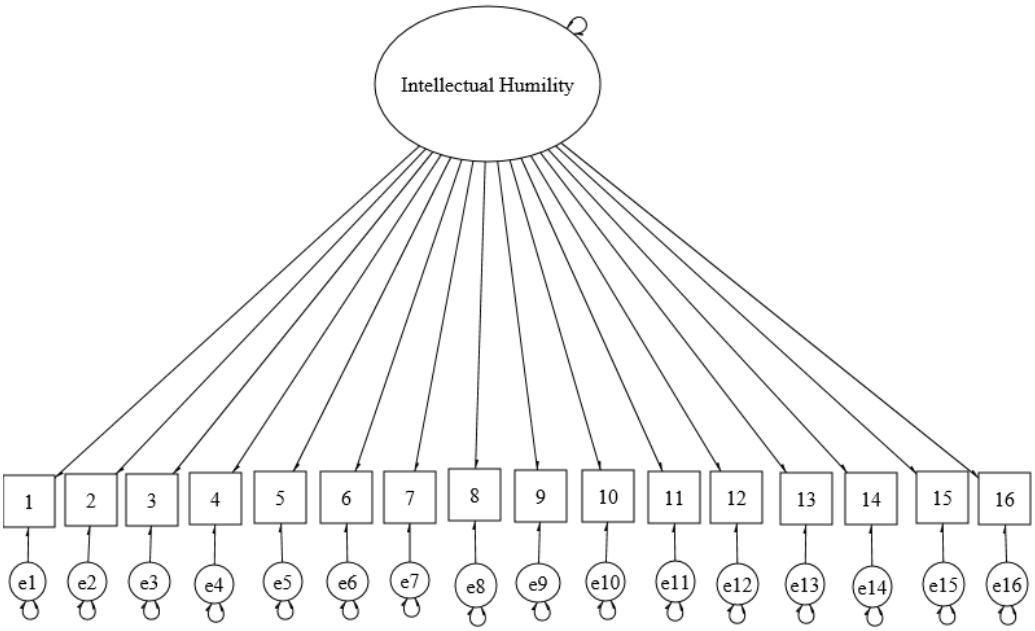

Figure 1. *One-factor model of intellectual humility.*

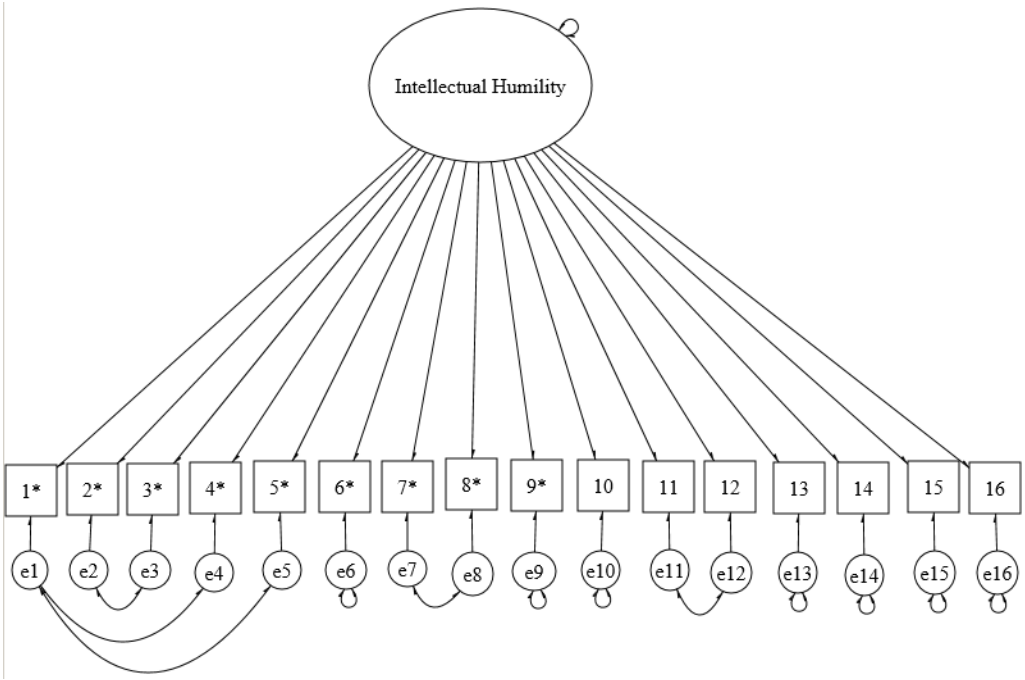

Figure 2. *One-factor model of intellectual humility with correlated residuals.*

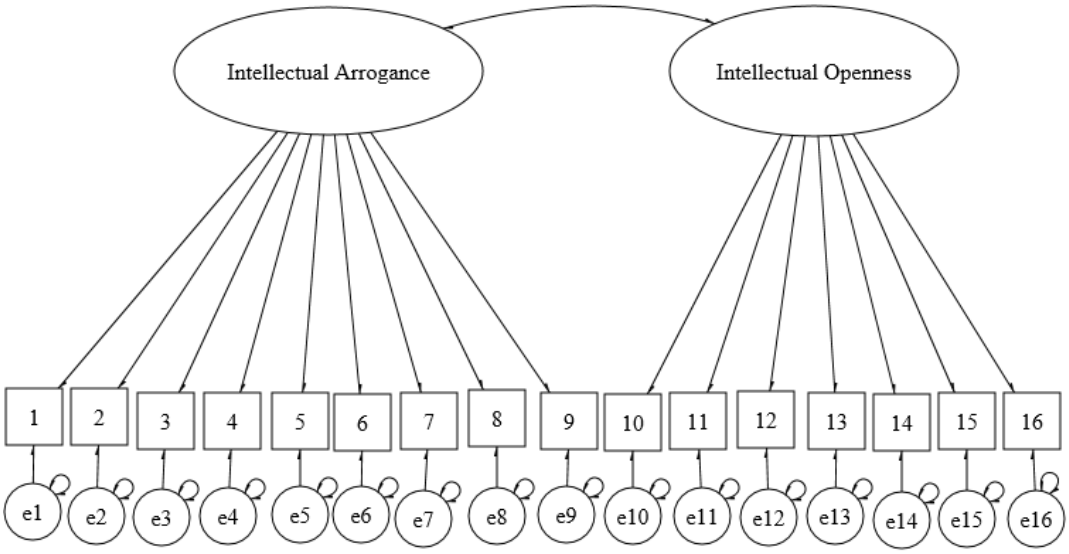

Figure 3. *Two-factor model for intellectual humility with factors of intellectual arrogance and intellectual openness.*

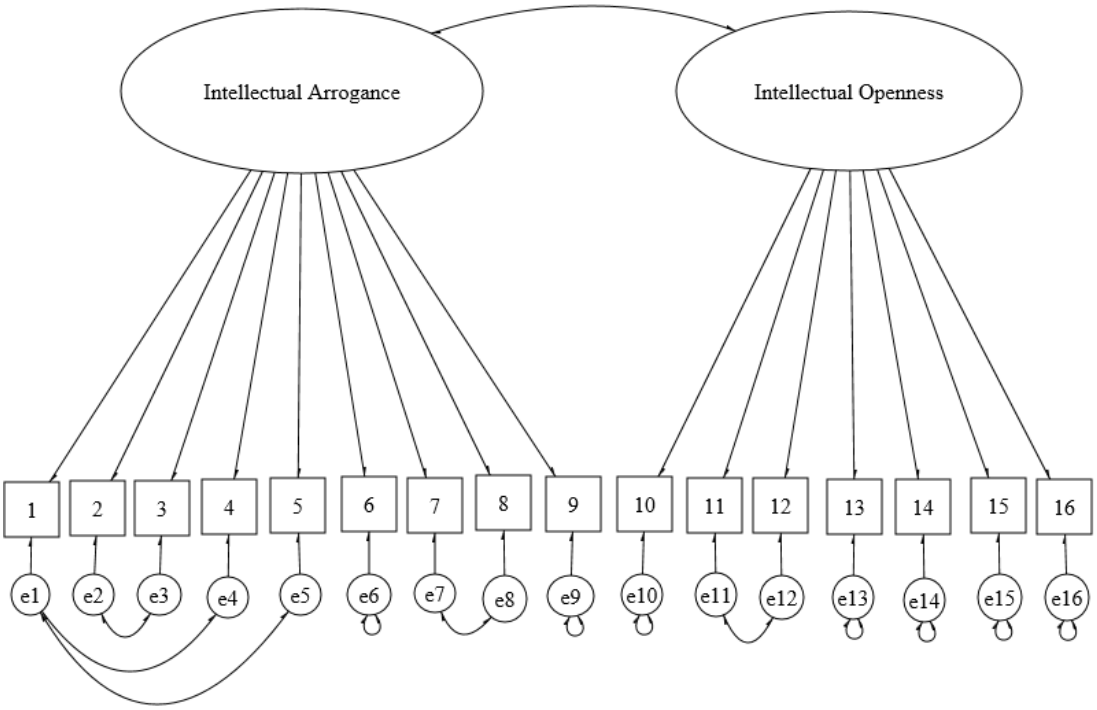

Figure 4. *Two-factor model for intellectual humility with factors of intellectual arrogance and intellectual openness with correlated residuals.*

Intellectual Humility, Arrogance, and Openness: Investigating the Psychometric Properties of a Self-Report Measure of Intellectual Virtue

The purpose of this study is to work toward developing a change sensitive measure of intellectual humility (IH). The conceptual grounding for this study begins in the intellectual virtues literature. From Aristotle to Aquinas, intellectual virtues have long been a point of interest for philosophers. Psychologists have engaged with the topic of intellectual virtues as well (e.g., Baltes, 1993; Sternberg, 1998; Thomas, Bangen, Ardelt, & Jeste, 2017). Baehr (2013) has argued that intellectual virtues represent an important dimension of character, as they are "the personal qualities or characteristics of a lifelong learner" (p. 249). Such qualities include curiosity and inquisitiveness, which help people to engage intellectually with others and with their environments over the life-course. To approach the empirical goal of this study, and assess the role of intellectual virtues that Baehr specifies, developmental scientists may study the role that specific virtues play in positive functioning throughout the course of human development.

IH is an example of one such intellectual virtue that may have an important role in what Baehr terms "lifelong learning," and, as such, it is the focus of this research. Roberts and Wood (2003) define IH as the accurate self-perception of one's knowledge in relation to others (i.e., neither prideful nor self-deprecating) and, as well, as an ability to believe information in accordance with the evidence provided. Similarly, McElroy et al. (2014) note that IH allows a person to be open to the ideas of others and to possess self-awareness of one's

arrogance, thereby enabling a person to offer and receive ideas with little or no offense.

In the current study, as a step towards the larger goal of establishing a change-sensitive measure of IH, I first conducted a literature review of preexisting measures of IH. After selecting a measure that seemed to be the best and most holistic representation of the construct—the Intellectual Humility Scale (IHS; McElroy et al., 2014)—I conducted various analyses to verify the factor structure of the measure, after translating it from a rater-report measure into a self-report measure. In addition to verifying the factor structure of the IHS, I conducted additional testing to assess if gender invariance for the scale could be established. This step was taken to ensure that the scale would be equally useful and "impartial," regardless of gender. To rationalize this research I discuss IH research, its current state, and what is still unknown in understanding this intellectual virtue. I then highlight the importance of integrating the extant work on IH with a developmental understanding of intellectual virtues, within the scope of life-long learning, and how the present study will enhance future IH research through the use of a developmental framework.

IH as a Psychological Construct

IH is knowing *and* accepting both what one contributes and cannot contribute intellectually in a given situation. IH implies that there are things that cannot be known, and that those who embody IH are virtuous for having such insight. A relatively small set of empirical studies have established positive associations between IH and various desirable attributes, such as self-regulation

(Dwiwardani et al., 2014), cognitive flexibility and openness (Jarvinen & Paulus, 2016; McElroy et al., 2014), and general humility (Davis et al., 2015). These empirical studies augment Baehr's (2013) claim of the contribution of intellectual virtue, in this case IH, to life-long learning, by highlighting the various positive attributes associated with such an intellectual virtue.

Despite the existence of these empirical connections between IH and other positive character and intellectual attributes, there are two points to emphasize as researchers move forward in understanding and assessing this construct: 1. there is sparse research about IH; and 2. extant IH research has not taken a developmental perspective. Over time, the first point will likely be addressed, given the current interest in IH within psychology (e.g., Davis et al., 2015; McElroy et al., 2015). The second point, however, is more of a challenge.

Theories of character development are still being debated within contemporary developmental science (e.g., Lerner & Callina, 2014). As Lerner and Callina (2014) note, research on character in psychology has often involved assessing traits, which are thought to be innate and immutable features of an individual's personality (e.g., McCrae et al., 2000). Research within social, evolutionary, and positive psychologies that focuses on traits (e.g., Davis et al., 2015; McElroy et al., 2014; Samuelson, Church, Jarvinen, & Paulus, 2013) has reduced character virtues to within-individual constructs, and therefore have not accounted for contextual influences. A contemporary, developmental approach to understanding character virtues rejects the trait conception of character and, framed by a process-relational paradigm and relational developmental systems

(RDS) metamodel (Overton, 2015a), emphasizes the mutual embeddedness of multiple levels of the ecology of human development (Lerner & Callina, 2014).

Whereas other useful approaches to developmental theory and research exist outside of the process-relational paradigm and RDS metatheory, I have chosen these frameworks to guide my work. Although developmental approaches to character virtue assessment have begun (e.g., Callina et al., 2017), this work is still in its preliminary stages. Thus, key methodological foci of an RDS-based approach to studying the development of character virtues, that is, using assessments that are change sensitive and that appropriately assess both the individual and the context, remain largely unaddressed. This situation is also the case with the study of IH.

However, an RDS-based framework can be used to devise a theoretical model for IH development and, in turn, for its measurement. Samuelson et al. (2013) agree, pointing to the lack of developmental, life-span research in the study of IH. Hence, developmental scientists have an opportunity to add new information about IH and, as well, about a potential feedback loop within the relational developmental system between IH and the contexts in which it is embedded, manifested, and revealed.

IH within a Process-Relational Paradigm and the RDS Metatheory

Cartesian-Split, mechanistic metatheory (see Overton, 2015a) is the predominant conceptual framework used to study character (e.g., Credé, Tynan, & Harms, 2016; McCrae et al., 2000). This metatheory frames psychological approaches that reduce character to genetic mechanisms. The alternative to such a

metatheory is the process-relational paradigm. This conceptualization is, as Overton (2015b) notes, "committed to an ontology of change… [and] a relational holism" (p. 166), with a focus on one's agency within and across time and contexts, and the nonlinearity of development. This paradigm allows researchers to assess the development of an individual, environment, or indeed any construct, in a holistic manner, synthesizing all possible factors of influence into one's science. This paradigm provides a framework through which IH might be best assessed from an RDS-based lens.

RDS is a metatheoretical frame through which theories can be devised (Overton, 2015b). Using RDS metatheory as a frame, Lerner (2015) and Overton (2015a) note that development involves person⇔context relations; dichotomous, split concepts (e.g., nature v. nurture) are incomplete and reductionist alternatives to the holistic RDS metatheory. Within this relational developmental system, the person is agentic, self-creating, and influences his or her context, as the context simultaneously, although not necessarily equally, shapes and changes the individual. The individual and context can only be understood in their entirety in reference to one another, never as mutually exclusive entities; therefore, both are necessary and basic components of human development (Lerner, 2015). The functioning of this system creates within-person change and between-person differences in such change, and thus enables relative plasticity across time and place (Lerner, 2015).

Lerner and Callina (2014) have suggested that RDS-based theories would be useful to study character development, and the development of virtues more

specifically. From the RDS perspective, "character… is a multidimensional, relatively plastic feature of adaptive developmental regulations, of mutually beneficial individual⇔context relations" (Lerner & Callina, 2014, p. 333). This statement certainly pertains to IH. Within the RDS frame, research on IH could generate knowledge of associations between the person and the context, such as the presentation of new information, in various contexts, within various relationships, and with various epistemic peers, in addition to individual variables, such as self-regulation and cognitive flexibility. In addition, RDS-based theories could elucidate how these person⇔context relations are associated with broader levels of the ecology to influence development of IH in and across individuals.

Developmental Considerations about IH

Establishing the need for a process understanding of IH, as it is embedded in a developmental context and in relation to the concept of life-long learning, as proposed by Baehr (2013), will require theory-predicated research. I will eventually propose and test a developmental model of IH and its relations to the context within which it develops. A key initial step is to establish a new measure, or empirically validate a preexisting, psychometrically sound measure, of IH. To conduct this program of research to developmentally investigate IH, I will use a data set associated with an ongoing longitudinal study involving Tufts University and the United States Military Academy (USMA), termed Project Arete.

Arete is Greek for *moral virtue* or *moral excellence*. Project Arete is a five-year longitudinal, multi-method and multi-rater study. The primary goal of the project is to assess the developmental processes involved in West Point's

mission, that is, to graduate leaders of character who will become officers in the United States Army (*Gold Book*, 2015). The purpose of the project is to better understand character virtue development at USMA, in order to describe, explain, and optimize this development for cadets, as well as to assess what could be applicable to other contexts of higher education and student experiences. Although, at this writing, data collection has only recently begun, there are data gathered about IH.

The Intellectual Humility Scale (IHS; McElroy et al., 2014) was included in the preliminary data collection of the project. This measure is one of the few existing ones of IH. As mentioned above, IH is a recent concept in the psychological literature, far behind its conceptualization in philosophy and theology (McElroy et al., 2014). Krumrei-Mancuso (2016) reported that "only six empirical articles" (p. 3) existed pertaining to the topic of intellectual humility. At this writing, only about a dozen or so exist.

McElroy et al. (2014) were the first of only a few research teams to develop a measure of intellectual humility, labeled the Intellectual Humility Scale (IHS). The original measure was designed as part of a larger study of trust of religious leaders, using a majority-women (74%) sample from Amazon Mechanical Turk, a popular platform for recruiting participants. The IHS was devised by winnowing down items from 60 original items by way of an exploratory factor analysis. Eventually, a 16-item "other-report" measure was developed. In a second study conducted by McElroy et al. (2014), the factor structure of the measure was verified using confirmatory factor analysis (CFA). In

this study, a sample (55% women) was again recruited from Amazon Mechanical Turk. In another study in the same publication, McElroy and colleagues used the IHS with another sample (71% women), composed of college students from a public university. In each study, the respective CFAs resulted in good model fit, and with estimates of internal consistency reliability ranging from .80 to .96 across studies.

Within the IHS measure, McElroy and colleagues (2014) assessed two subscales of IH, which they hypothesized to be the major components of IH: intellectual openness (IO) and intellectual arrogance (IA). IO is defined as an openness to others' ideas, whereas IA is defined as intellectual impatience or invalidation of others. With these two subscales, the authors pursued a similar investigation of the factor structure as they did with the holistic measure of IH. The researchers conducted both an EFA and a subsequent CFA in order to investigate whether a two-factor model of IA and IO, rather than the general factor of IH, provided a better fit to the data. The two-factor CFA model provided a better fit than the one-factor model. Although the researchers conducted rigorous testing of the best model fit for the IHS, McElroy et al. (2014) did not conduct tests of measurement invariance for gender, despite relying heavily on samples of women.

Gender differences in moral development have long been a topic of discussion in the field of developmental science. To use a well-known example, Kohlberg and Kramer (1969) used specific measures of moral development, ones that eventuated in evidence that women may not as morally developed as men.

Gilligan (1982) then contested this evidence, postulating that women approached moral situations in a different way than men; she argued that measurement of moral behavior and development was different for men and women (Jaffe & Hyde, 2000). In addition, personality traits (e.g., Weisberg, DeYoung, & Hirsh, 2011) and other character attributes (Park & Peterson, 2006) have shown notable gender differences across various studies. For these reasons, it is important to understand if measurement tools used to assess character and intellectual virtues are invariant across gender and, perhaps as well, other variables (such as age, race and SES).

The IHS has not only been used by McElroy and colleagues (2014), but has appeared in studies since the original measurement article was published. For example, Davis and colleagues (2015), in a study showing the differentiation of general humility from intellectual humility, used the IHS, but used the scale as a self-report measure. In other words, the language of the scale was not altered, but the target of the rater was himself or herself, and not another individual. This approach is one way of using the IHS as a self-report, but another method of self-report would be to change the wording of the rating items to be focused on the self.

Despite the promising results of the "other-report" IHS, it is important to have a reliable and valid self-report measure of IH. In a report comparing self-report and consensus ratings of IH, Meagher, Leman, Bias, Latendresse, and Rowatt (2015) showed differences in personal (self-report) and relational (other-report) IH assessment. There was greater consensus between self and others on

items that were associated with more overt behaviors, and this greater consensus was only achieved when those doing the relational assessments had long-term contact with the target. Having a self-report measure of a construct can be useful to get insight into attributes that are otherwise difficult to evaluate without behavioral indicators, because other people may be less accurate in evaluating or even having an awareness of private or less visible attributes (Vazire, 2010). However, Meagher et al. (2015) noted that there exists little empirical validation of self-report measures of IH.

The Current Study

Given the lack of intellectual humility measures in the empirical literature, and that the primary available measure requires others to report about a target participant, I adapted the IHS to be a self-report measure. Although one set of researchers (Davis et al., 2015) used the IHS as a self-report measure, they did so by changing the target of the IHS other-report to the self (e.g., rating the self on an item such as "Acts like a know-it-all") rather than changing the wording of the items to be self-focused (e.g., by using "I" in the statements). In addition, Davis and colleagues (2015) did not provide any psychometric analyses of their target-as-self method. The analyses proposed here are an assessment of the measurement properties of the IHS as a self-report measure by using self-focused items, and with a much larger and more diverse sample than the McElroy et al. (2014) samples. I compared the fit of a single-factor (i.e., IH) model and a two-factor model (i.e., subscales of IO and IA). In addition, I tested the best model, from the aforementioned tests of model fit, for measurement invariance across gender.

Method

Data Source

The data were derived from Project Arete, a larger study of the development of character virtues among cadets at the United States Military Academy (USMA), both during their time as cadets, and into their careers as Army officers. This research is a collaborative effort between Tufts University and USMA. The overall study is a longitudinal, 5-year cohort sequential, mixed-method investigation, seeking to understand the individual character strengths of the cadets, as well as how these strengths may be enhanced by the West Point context and experience. Data came from the first wave of data collection with new cadets before they entered into basic training.

Participants

For the current analyses, 1,257 first-year USMA cadets (78% men; 64.8% White; 12.7% Black; 9.6% Asian; 9.5% Hispanic) took part in a self-report survey of several character virtues, one of which was IH. Age was not reported on the survey. However, the average age of incoming cadets tends to range from 18-19, unless cadets have prior military service. Gender was reported as a binary variable. The low representation of women and racial minorities in the sample of cadets is consistent with the history of West Point, and the general military populations, which, until recently, had been exclusively male, and majority white. The implications of these imbalances will be addressed in the Discussion section.

Measure

Intellectual humility (IH). In order to assess intellectual humility, I modified the preexisting "other-report" measure, the Intellectual Humility Scale (IHS; McElroy et al., 2014), into a self-report measure. For example, the item "Has little patience for others' beliefs" was recast as "I have little patience for others' beliefs." The 16-item measure consists of two subscales, Intellectual Arrogance (IA; 9 items; e.g., "I act like a know-it-all") and Intellectual Openness (IO; 7 items; e.g., "I enjoy diverse perspectives"). Table 1 presents the stems for all items. The items have five response options, from 1= *strongly disagree* through 5 = *strongly agree.* Accordingly, when the subscales are considered separately, higher scores on the IO items indicate more IH, and higher scores on the IA items indicate less IH. When considering IH holistically, the IA items are reverse-coded, and higher scores indicate higher overall intellectual humility.

Procedure

Recruitment took place in the summer of 2016, from the incoming class of cadets at the United States Military Academy at West Point. Cadets completed a multiple-choice paper survey and a Scantron scoring sheet during a one-hour testing period during their pre-Basic Training New Cadet Testing Period, which took place between June 27 and 29, 2016. Cadets took the survey grouped by their company (A-H), two companies per session, across the three days, with two testing sessions on the first day. After the testing session, responses from the Scantron sheets were then entered into the Statistical Package for the Social

Sciences version 22 (SPSS 22; IBM Corp., 2013), where the data were de-identified.

Analysis Plan

In order to assess the factor structure of the self-report IHS, I conducted two types of CFAs using the statistical package Mplus 7.4 (Muthén & Muthén, 2016). The first CFA was a one-factor model of IH, and the second was a two-factor model of the IHS, with IA and IO as the two latent factors. I then compared the results of these CFAs to see which one fit the data better. The best factor structure for the construct was assessed using recommended cutoff values (see Results) for three statistical tests of model fit, the Confirmatory Fit Index (CFI) (a goodness-of-fit indicator) and, in turn, two badness of fit indicators: the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR). Once I established which factor structure best fit the data, I used the best-fitting model to conduct invariance testing of gender.

Invariance testing requires little-to-no change in the CFI (cutoff of $\Delta CFI \leq .01$) between three modifications of the factor structure, split by gender: 1. configural invariance, which tests whether the two groups display the same pattern of factor loadings; 2. loading invariance, which tests whether the two groups display equal pattern coefficients and 3. intercept invariance, which tests whether the two groups display equal intercept values given the same level of the latent factor. As a general rule, the measure is shown to be invariant if, (a) the $\Delta CFI \leq .01$ from configural invariance to loading invariance; and (a) the $\Delta CFI \leq .01$ from loading invariance to intercept invariance (Kline, 2016). In other words,

invariance testing can show that the measurement of a specific latent variable, that is, IH, can be used reliably across gender, and that mean-level comparisons of the construct are appropriate.

Results

I will now present the results of the analyses, assessing the factor structure and the invariance of the measure of IH. However, before these primary analyses were done, initial preliminary analyses were undertaken.

Preliminary Analyses

Preliminary analyses consisted of three steps: assessment of normality, assessment of missing data, and assessment of item-level correlations.

Normality. Intellectual humility scores were fairly normally distributed (Field, 2013), albeit with a mean score almost one point higher than the scale middle of 3 (M = 3.93). Whereas the items on the IO subscale were fairly normally distributed, the IA items were skewed slightly positively, which could be a function of the potential negative connotation attached to the IA items. I examined the data and determined that the amount of skewness was not large enough to invalidate using the items as continuous indicators, especially given the large sample size (Tanaka, 1987). Table 2 provides the descriptive statistics for each item in the measure.

Missing data. Missingness at the item level ranged from .0016% (2 cases) to .0088% (11 cases). Because of this very small proportion of missingness, missing items were handled using full information maximum likelihood (FIML)

in MPlus, because these data were assumed to be missing at random (Enders & Bandalos, 2001).

Item-level correlations. As the items were designed to be part of the same construct, there should be significant correlations among the items. For the sake of parsimony, however, items should not be too highly correlated, as that could imply redundancy. The cut-off used for evaluating potential item redundancy was $r = .70$ (Kleinbaum, Kupper, Muller, & Nizam, 1998). No pairs of the items were correlated above this threshold. Therefore, all items were retained in the CFA. Table 3 presents the item-level correlation matrix. I turn, then, to the primary analyses of the present study.

Confirmatory Factor Analyses

The purpose of these analyses was threefold. The first purpose was to assess how the IHS performed when used as a self-report measure. The second was to assess the factor structure when using a sample from the population different from the one used to validate the original "other" report measure (McElroy et al., 2014). The third purpose was to compare one-factor and two-factor representations of the factor structure to determine which provided the best fit. The models were estimated using maximum likelihood estimation.

Single-factor structure. For the single-latent factor structure of the IHS, all 16 items were specified to load on to the latent factor of IH. This model (Model 1; see Figure 1) was specified using the marker variable method and included no correlated errors. Table 4 presents unstandardized factor loadings of Model 1. As indicated by Hu and Bentler (1999), the following values tend to

suggest good fit: a CFI of about .95, an RMSEA of .06 or less, and an SRMR of .08 or less. In the case of this one-factor model, the model showed an overall lack of good fit, with some indices indicating better fit than others (X^2 (104) = 1663.894, p <.001, *CFI* = .669, *RMSEA* = .110, *SRMR* = .09). Some items loaded onto the latent factor better than others, with standardized factor loadings ranging from .27 to .66.

As a result of these poor fit indices, modification indices were requested in the analysis. Although Jackson, Gillaspy Jr, and Purc-Stephenson (2009), encourage sparse usage of post-hoc modifications to model fit, the newness of this construct suggested at least some additional data analysis for model improvement. Modification indices suggested correlated residuals between five pairs of items: 1 and 4, 1 and 5, 2 and 3, 7 and 8, and 11 and 12 (see Table 1 for item stems). Pairs of items that had suggested modification indices greater than 100.00 also tended to have inter-item correlations that were greater than r = .35. These modifications indices and correlations were indicators of empirical connections within the data set.

In addition to the post-hoc modification indices indicated by the data, there was a theoretical reason for correlating the item residuals for the respective pairs of items. Each of the item pairs with correlated residuals had very similar affective components to the item (e.g., Items 1 and 5 refer to anger; Items 2 and 3 refer to need to triumph in an argument). Therefore, I specified a modified one-factor model with correlated residuals among these five pairs of items. This modified specification displayed much improved fit, X^2 (99) = 1054.625, p < .001,

CFI = .797, *RMSEA* = .088, *SRMR* = .079. The Chi-square difference between the original and modified specifications was also significant, ΔX^2 (5) = 609.269, $p <$ .001. The range of factor loadings was even greater for this model specification, ranging from .23-.68. Table 5 presents unstandardized factor loadings.

Two-factor structure. In order to assess whether a two-factor structure would improve model fit from the one-factor structure, a two-factor model was specified using the item loadings for IA and IO as outlined in the original IHS measurement paper (McElroy et al., 2014). The first nine items were loaded on to IA, and the latter seven items were loaded on to IO, using the model specifications of the original measurement paper (McElroy et al., 2014). To assess modifications and potential improvements to model fit using a two-factor model, a different two-factor model was specified: one using the marker variable method with uncorrelated residuals (Model 3; see Figure 3), and one with the correlated residuals identical to Model 2 (Model 4; see Figure 4). In addition, for both models, the two latent factors were correlated (as is the default). Accordingly the two-factor structure was the only difference in model specification from the one-factor structure; thus the specifications were easier to compare. The factor loadings for Model 3 had a narrower range in comparison to Model 2. The standardized loadings for Model 3 ranged from .42 to .60 for IA, and from .41 to .75 for IO. See Tables 6-7 for unstandardized factor loadings for Models 3 and 4.

Using the recommended guidelines indicated by Hu and Bentler (1999), the fit indices for Model 3 indicated moderate fit for the data, X^2 (103) = 750.209, $p <$ .001, *CFI* = .861, *RMSEA* = .071, *SRMR* = .058. Model 4 indicated fairly

good fit for the data and was by far the best fit for the data of the four models specified in the current analyses, X^2 (98) = 512.847, $p < .001$, *CFI* = .911, *RMSEA* = .058, *SRMR* = .048. In addition, the fit indices for Model 4 were a significant improvement from both Model 2, ΔX^2 (1) = 541.778, $p < .001$ and Model 3, ΔX^2 (5) = 237.362, $p < .001$. The unstandardized factor loadings for Model 4 were all close to 1.00 for both latent factors, and all of the items loaded fairly similarly onto their respective latent factors (ranging from .86 to1.26 for IA, and .83 to 1.20 for IO). The standardized factor loadings ranged from .38 to .60 for IA, and from .42 to .71 for IO. The standardized residual correlations for the five pairs of items were moderate (*r* ranging from .13 to .26). The latent factor correlation was *r* = -.57, which indicates similarities and uniqueness between the two latent factors. This strength of correlation is ideal, given that IO and IA represent distinct constructs of the higher-order latent variable of IH.

Measurement Invariance

As indicated at the beginning of this paper, the original measurement studies for the IHS (McElroy et al., 2014) were conducted, at times, with a gender-imbalanced sample. Given that the sample used in the current study was also gender-imbalanced (there was a larger percentage of men), measurement invariance by gender was necessary to provide evidence that the latent construct of IH was being measured the same way across genders. Measurement invariance of the two-factor structure of the IHS, with the five correlated residuals (see above), was conducted using the demographic information provided by USMA

within the Office of Institutional Research, which used a binary measure of gender, Male or Female.

Overall, most factor loadings were similar across genders, and were thus similar to the overall loadings noted above (Table 8 presents standardized loadings by gender from the configural model). The two-factor structure of IH was invariant across genders (cutoff for $\Delta CFI \leq .01$), except for intercept invariance ($\Delta CFI = .012$), when all corresponding intercepts were estimated to be equal in both groups. Because the intercepts were not invariant across groups, the intercepts across the groups were examined for instances of dissimilarity. Item 12 ("I enjoy diverse perspectives") showed a significant difference in standardized intercept values between genders ($\Delta\beta_{012} = 0.646$), and was thus specified with unequal intercepts between the two models. Setting the intercept for Item 12 unequal across the two models resulted in a $\Delta CFI = .008$. That is, the structure showed overall measurement invariance across gender, when Item 12 was specified with unequal intercepts. Table 9 presents the findings for configural, loading, and intercept invariance.

The results of these analyses indicated that, out of four models specified, a model involving two components of the IHS, IA and IO (Model 4), was the best fit for the data. This model specified five pairs of correlated residuals in order to improve model fit. Using Model 4, the measure showed configural and loading invariance by gender when all items were included in the model specifications. When Item 12 was specified with unequal intercepts, IH showed full measurement invariance by gender.

Discussion

Intellectual humility (IH), and intellectual virtues in general, are character attributes that may contribute to life-long learning (Baehr, 2013) and, as such, should be conceptualized as developmental. The structure and function of IH may change across the life span. In order to developmentally assess such constructs, measures should be reliable, valid, and change-sensitive. Unfortunately, IH has yet to be developmentally assessed and, overall, measures of the construct lack empirical validation, especially in a self-report format (Meagher et al., 2015).

Therefore, the purpose of the present study was to investigate the psychometric properties of a self-report measure of IH. The current study provided evidence that a self-report adaptation of an IH measure, the IHS (McElroy et al., 2014), had sound psychometric properties, and could be used across genders. More specifically, the results of this study supported two conclusions: 1. The best factor structure for the self-report adaptation of the IHS was a two-factor structure with correlated errors, specified using the subscales as latent factors of intellectual arrogance (IA) and intellectual openness (IO); and 2. with a minor modification, IH showed measurement invariance by gender using the two-factor model of IH as specified in Conclusion 1.

The results of the CFAs are perhaps unsurprising, given that the self-report IHS is derived from a psychometrically sound, preexisting measure. Nevertheless, it is interesting that psychometric quality was maintained across a marked change in format, from “other report” to “self-report.” Regardless of the well-fitting factor structure of the two-factor model of IH, the invariance testing by gender

was a positive, and previously uninvestigated, outcome of the current analyses. Measurement invariance of IH by gender, as measured by a self-report version of the IHS (when Item 12 is specified with unequal intercepts), indicates, at least within the USMA population, that men and women are answering the self-report IHS in similar ways. This result means that researchers can utilize this tool in the future, and do not necessarily need to be concerned about gender being a complicating variable for participant responses. Future studies using a USMA sample should test for measurement invariance by gender, though, if mean comparisons are going to be made.

By establishing the foundation for a psychometrically-sound self-report measure of IH, researchers can investigate how self-report measures of IH compare and contrast to other formats for measuring IH (e.g., other-reports; Meagher et al., 2015; coding of IH-relevant interviews). Such triangulation may help elucidate how self-report measures of IH can be improved. Possessing psychometrically-sound measures of IH can be a first step in establishing the change sensitivity of such measures and, as such, could be used in future research about the role of IH in life-long learning and in individual⇔context relations across time and place.

Although there is potential, therefore, in the future use of the IH self-report measure developed in the present research, there are important limitations of the results that should be noted. Some items had residuals that, when correlated, significantly improved the model fit of the specified CFA models. A potential justification for correlating the residuals was presented in the Results

section, that is, that these pairs of items have similar affective content. Whereas this correspondence in content may be true, these high correlations could also indicate item redundancy. Future studies should investigate whether these items, and others, add significant value to model fit for measuring these latent variables, or if these items are indeed redundant and should be eliminated. Similarly, the results of tests of measurement invariance of IH by gender indicated that removing a specific item—in this case Item 12—from the scale improved the reliability and validity of measuring IH. Item 12 may uniquely be a point of interest for women at West Point because they are selecting to be in an environment with "diverse perspectives," that is, where they are the minority; as such, the unequal intercepts for this item may not apply to other samples in future studies. In future tests of the use of the present measure of IH with other samples, researchers might test if eliminating Item 12 enhances model fit and/or if the removal of this item in the present invariance analysis was needed only for improving fit with this sample.

By moving toward a more accurate measure of the latent construct of IH, researchers can also begin to investigate Baehr's (2013) claim that intellectual virtues are indeed characteristics of life-long learning. Using an RDS metatheoretical framework (Overton, 2015a), developmental scientists can theorize how the individual and his or her IH co-act with his or her context. Once individual patterns of individual⇔contexts relations are established for IH, the intellectual virtue's role in life-long learning can be investigated. The present

study is, then, only a first step in eventually understanding the role of IH in life-long learning.

Another key limitation of this research involves the population that was surveyed. West Point cadets were the only participants in this study. As such, the results here may only be generalizable to the West Point population. Such generalizability should be tested with similar populations (e.g., from other service academies). In addition, the racial distribution of the sample was mostly White. Thus, future research should include greater diversity of participants and focus on measurement invariance as it pertains to these other groups.

The self-report measure of the IHS is limited by the ability of an individual to accurately and truthfully report on his or her IH, as well as by the item sentence structure and response scale. IH is a somewhat socially desirable construct, which can potentially skew accuracy of self-reports. For this reason, outside raters may be better at detecting IH (Vazire, 2010). In addition, the IHS contains several strongly valenced items, which might lead some respondents with a bias towards agreeing with strongly worded items to choose more extreme response options (Brown, 1965; and in the current study, either a *1* or a *5*; Furnham, 1986); however, as indicated by the nature of the sample distribution, I did not encounter such bias with the current sample.

In sum, in order to both improve the self-report IHS measure and best understand IH as a holistic construct, the IHS must be triangulated with other methodologies and measures of IH. Studies of IH should continue to investigate the differences, benefits, and limitations of self-report and "other-report"

measures of IH, both quantitatively and qualitatively. By establishing tools that are appropriate across diverse groups and diverse raters, researchers can come closer to understanding the role of IH in the lives of young adults.

Conclusions

By integrating these findings with future studies of IH, researchers can begin to develop a model sensitive to changing individuals and contexts, and move towards an idiographic understanding that Bornstein (2017) illustrates in regard to his discussion of the specificity principle. The principle highlights the importance of understanding the process of specific attributes, in specific contexts, with specific individuals, who have specific experiences, demographics, etc. By using results across studies of IH with varying groups of individuals (e.g., cadets, professors, and barbers) at varying ages (e.g., college-aged, infants, and retirees) in varying contexts (e.g., a military academy, one's home, and one's workplace), developmental scientists can better describe how IH will manifest, as context, individual, and experiences change within and between individuals and groups.

Although the present analyses are preliminary, not yet longitudinal, and not completely encompassing the entirety of IH across individuals and contexts, the present research may be an impetus for such future research. The analyses undertaken in this research may be a basis for building a developmental model for IH, one that shows the potential embeddedness of IH in life-long learning (Baehr, 2013).

References

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

Baltes, P. B. (1993). The aging mind: Potential and limits. *The Gerontologist*, *33*(5), 580-594.

Bornstein, M. H. (2017). The specificity principle in acculturation science. *Perspectives in Psychological Science, 12*(1), 3-45.

Brown, R. (1965). *Social psychology*. New York: The Free Press.

Callina, K. S., Ryan, D., Murray, E. D., Colby, A., Damon, W., Matthews, M., & Lerner, R. M. (2017). Developing leaders of character at the United States Military Academy: A relational developmental systems analysis. *Journal of College and Character*, *18*(1), 9-27.

Credé, M., Tynan, M. C., & Harms, P. D. (2016). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology.*

Davis, D. E., Rice, K., McElroy, S., DeBlaere, C., Choe, E., Van Tongeren, D. R., & Hook, J. N. (2015). Distinguishing intellectual humility and general humility. *The Journal of Positive Psychology*, 1-10.

Dwiwardani, C., Hill, P. C., Bollinger, R. A., Marks, L. E., Steele, J. R., Doolin, H. N., Wood, S.L., Hook, J. N., & Davis, D. E. (2014). Virtues develop from a secure base: Attachment and resilience as predictors of humility, gratitude, and forgiveness. *Journal of Psychology and Theology*, *42*(1), 83-90.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430-457.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and individual differences*, *7*(3), 385-400.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55.

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods*, *14*(1), 6.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, *126*(5), 703-726.

Jarvinen, M. J., & Paulus, T. B. (2016). Attachment and cognitive openness: Emotional underpinnings of intellectual humility. *The Journal of Positive Psychology*, 1-13.

Kleinbaum, D.G., Kupper, L.L., Muller K. E., & Nizam, A. *(1998). Applied regression analysis and other multivariable methods.* CA: Brooks/Cole Publishing Company.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Publications.

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. Human Development, 12, 93-120.

Krumrei-Mancuso, E. J. (2016). Intellectual humility and prosocial values: Direct and mediated effects. *The Journal of Positive Psychology*, 1-16.

Lerner, R. M. (2015). Preface. In W. F. Overton & P. C. Molenaar (Eds.), *Theory and Method* Volume 1 of the *Handbook of Child Psychology and Developmental Science* (7th ed.)(pp. xv-xxi). Hoboken: Wiley.

Lerner, R. M. & Callina, K. (2014). The study of character development: Towards tests of a relational developmental systems model. *Human Development, 57*(6), 322-346.

McCrae, R.R., Costa, P.T. Jr., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M.D., Sanz, J., Sanchez-Bernardos, M.L., Kusdil, M.E., Woodfield, R., Saunders, P.R., & Smith, P.B. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology, 78*, 173-186.

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington Jr, E. L., & Tongeren, D. R. V. (2014). Intellectual humility: Scale development and theoretical elaborations in the context of religious leadership. *Journal of Psychology and Theology, 42*(1), 19-30.

Meagher, B.R., Leman, J.C., Bias, J.P., Latendresse, S.J., & Rowatt, W.C. (2015). Contrasting self-report and consensus ratings of intellectual humility and arrogance. *Journal of Research in Personality, 58*, 35-45.

Muthén, L. K., & Muthén, B. O. (1998-2016). *Mplus user's guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Overton, W. F. (2015a). Process and relational-developmental systems. In W. F. Overton & P. C.M. Molenaar (Eds.), *Theory and method. Volume 1 of the handbook of child psychology and developmental science* (7th ed., pp. 9–62). Hoboken, NJ: Wiley.

Overton, W. F. (2015b). Taking conceptual analyses seriously. *Research in Human Development*, *12*(3-4), 163-171.

Park, N., & Peterson, C. (2006). Moral competence and character strengths among adolescents: The development and validation of the Values in Action Inventory of Strengths for Youth. *Journal of Adolescence*, *29*(6), 891-909.

Roberts, R. C., & Wood, W. J. (2003). Humility and epistemic goods. In L. Zagzebski & M. DePaul (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology (*pp. 257-279). New York: Oxford University Press.

Samuelson, P. L., Church, I. M., Jarvinen, M., & Paulus, T. (2013). The Science of Intellectual Humility White Paper. Unpublished manuscript.

Sternberg, R. J. (1998). A balance theory of wisdom. *Review of General Psychology*, *2*(4), 347.

Thomas, M. L., Bangen, K. J., Ardelt, M., & Jeste, D. V. (2017). Development of a 12-ttem abbreviated three-dimensional wisdom scale (3D-WS-12): Item selection and psychometric properties. *Assessment*, *24*(1), 71-82.

United States Military Academy. (2015). *Character Program (Gold Book)*. West Point, New York. Retrieved 23 May 2016 from: http://www.usma.edu/scpme/SiteAssets/SitePages/Gold%20Book/Gold%20Book.pdf.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281-300.

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, *2*, 178.