J Youth Adolescence (2017) 46:1216–1237
DOI 10.1007/s10964-017-0660-1

EMPIRICAL RESEARCH

Modeling Pathways of Character Development across the First Three Decades of Life: An Application of Integrative Data Analysis Techniques to Understanding the Development of Hopeful Future Expectations

Kristina Schmid Callina[1] · **Sara K. Johnson**[1,2] · **Jonathan M. Tirrell**[1] · **Milena Batanova**[1] · **Michelle B. Weiner**[3] · **Richard M. Lerner**[1]

Received: 22 February 2017 / Accepted: 8 March 2017 / Published online: 22 March 2017

Abstract There were two purposes of the present research: first, to add to scholarship about a key character virtue, hopeful future expectations; and second, to demonstrate a recent innovation in longitudinal methodology that may be especially useful in enhancing the understanding of the developmental course of hopeful future expectations and other character virtues that have been the focus of recent scholarship in youth development. Burgeoning interest in character development has led to a proliferation of short-term, longitudinal studies on character. These data sets are sometimes limited in their ability to model character development trajectories due to low power or relatively brief time spans assessed. However, the *integrative data analysis* approach allows researchers to pool raw data across studies in order to fit one model to an aggregated data set. The purpose of this article is to demonstrate the promises and challenges of this new tool for modeling character development. We used data from four studies evaluating youth character strengths in different settings to fit latent growth curve models of *hopeful future expectations* from participants aged 7 through 26 years. We describe the analytic strategy for pooling the data and modeling the growth curves. Implications for future research are discussed in regard to the advantages of integrative data analysis. Finally, we discuss issues researchers should consider when applying these techniques in their own work.

Keywords Character virtue development · Integrative data analysis · Hopeful future expectations · Positive youth development

Introduction

There were two purposes of the present research. We sought to add to scholarship about a key character virtue—hopeful future expectations—and to demonstrate a recent innovation in longitudinal methodology that may be especially useful in enhancing the understanding of the developmental course of this and other character virtues that have been the focus of recent scholarship in youth development (e.g., Lerner and Callina 2014; Nucci in press). As exemplified by the growing literature about the development of hopeful future expectations, and its role as a moderator of processes of youth-context relations involved in the development of thriving during the adolescent years (e.g., Callina et al. 2014a; Schmid and Lopez 2011; Schmid et al. 2011a, b), the burgeoning interest in character development and character education across multiple domains of inquiry has led to a proliferation of data about positive youth development in various contexts and across the first three decades of life. On their own, these data may be limited in their ability to model trajectories of character development due to low power or relatively brief time spans in which to model change. However, in their "ongoing quest to build a cumulative psychological science," Curran (2009, p. 77) and

✉ Kristina Schmid Callina
kristina.callina@tufts.edu

[1] Institute for Applied Research in Youth Development, Tufts University, 26 Winthrop Street, Medford, MA 02155, USA

[2] Eliot-Pearson Department of Child Study and Human Development, Tufts University, 26 Winthrop Street, Medford, MA 02155, USA

[3] Boston College, Chestnut Hill, MA 02467, USA

colleagues propose statistical techniques that may be useful in building a cumulative science of character virtue development. In this article, we describe these techniques and their application to studying hopeful future expectations.

Hopeful future expectations are one construct of many in the array of psychosocial variables and processes that comprise positive, future-oriented cognitions and emotions, including Snyder's Hope Scale (Lopez et al. 2009; Snyder et al. 1991), optimism (Carver and Scheier 2002; Seligman 1991), possible selves (Markus and Nurius 1986), and future orientation (Seginer 2009; Steinberg et al. 2009). Such variables are assets in the character virtue development of young people in that they help organize and energize behavior toward a productive adulthood. For example, hopeful future expectations was shown to moderate the role of intentional self-regulation in promoting the "Five Cs" of positive youth development (Confidence, Competence, Character, Connection, and Caring; Schmid et al. 2011a, b).

As is the case with much of the extant data on positive youth development, different studies on hopeful future expectations have appraised only small portions of the life span, giving researchers and practitioners merely snapshots of the development of this key variable across childhood, adolescence, and early adulthood. We believe the theoretical importance attributed to this variable should be matched, at least, by an empirical description of the development of hopeful future expectations in the first three decades. Historically, such a description would require a data set too costly for most researchers to pursue.

Fortunately, emerging analytic techniques offer a promising way to link multiple independent samples. These techniques have the potential to address the theory-data discrepancy that exists in regard to hopeful future expectations and other character virtues (Card in press). Meta-analysis is one such technique that may bring new, integrative information to bear on the development of character and its role in thriving from childhood through early adulthood. In meta-analysis, parameter estimates are pooled from existing studies (Cooper and Patall 2009; Cumming 2013). Curran, Hussong, and colleagues (Bainter and Curran 2015; Curran et al. 2008; Curran and Hussong 2009) describe another approach, termed *integrative data analysis* (IDA), which allows researchers to pool raw data from existing studies in order to fit one model to the aggregated data. We propose that this analytic technique is a promising new tool for modeling character development. Accordingly, one purpose of this article is to provide a demonstration of the uses and limitations of the IDA approach and to highlight its potential utility for pooling data pertinent to character development.

Since 2012, the Institute for Applied Research in Youth Development has launched several research projects that focus on character as a key component of positive youth development. One goal of this research program is to assess positive character development (character virtue development) across the first three decades of life, in order to better understand the individual strengths and contextual assets that are associated with character virtue development from childhood through early adulthood, as well as the potential implications of character virtue development for human flourishing. Following the model of character development presented by Lerner and Callina (2014) and others (Berkowitz 2011; Lapsley and Narvaez 2006; Sokol et al. 2010), we define character attributes as the set of behaviors, attitudes, and emotions that allow an individual to "do the right thing" (morally and behaviorally) for self and others, and therefore to thrive in a specific context at a specific time (Nucci in press).

In sum, we focus on the character virtue of hopeful future expectations to both advance understanding of the course and importance of this attribute for positive development across adolescence and to assess the usefulness of IDA techniques to elucidate this development. To enact the two goals of this research, we were fortunate to be able to capitalize on the presence of several longitudinal data sets wherein measures of hopeful future expectations had good psychometric properties. It is useful to discuss in more detail the current knowledge base regarding hopeful future expectations and to then note how this study addresses the two purposes of the present research.

The Role of *Hopeful Future Expectations* in Character Virtue Development

The data sets pertinent to hopeful future expectations used in the present study are linked through a common theoretical framework known as *relational developmental systems theories* (e.g., Overton 2015). These theories point to mutually beneficial relations between the individual and context as the wellspring of thriving across the life span. Character virtues develop concomitantly with these mutually beneficial relations; that is, character virtues are the constellation of psychosocial behaviors and attitudes that both contribute to and result from positive and adaptive functioning within a particular context (Berkowitz 2012; Lerner and Callina 2014). Simply, character virtues allow the individual "to live well the life that is good for one to live" in his or her community (Lapsley and Narvaez 2006, p. 271).

Hopeful future expectations represent a unique facet of character virtue; indeed, hope is among the highest virtues in many religious and philosophical texts, and is often associated with faith and spirituality (Callina, Snow, & Murray, in press). From a developmental perspective, a young person's goals, expectations, and emotions about the future are powerful forces in shaping his or her life path

(Seligman et al. 2013). Adolescents' imagined futures, or *possible selves*, motivate present-day behaviors: many of the decisions that adolescents make, such as selecting friends, choosing extra-curricular activities, and deciding what to study, are reflections of their hoped-for future selves (Markus and Nurius 1986). Moreover, hopeful future expectations are important motivators for applying the goal-setting and management skills that allow individuals to attain their future goals and function adaptively (Callina et al. 2014b; Schmid and Lopez 2011). Thus, hopeful future expectations contribute to the mutually beneficial relations that support character virtue development across the life span.

As noted above, current research provides only snapshots into the development of this key variable. One study modeled hopeful future expectations among a national U.S. sample of youth in Grades 7 through 10 (Callina et al. 2014a). The study suggests that most young people have high or moderate levels of hopeful future expectations across these 4 years, with little fluctuation; however, some of the participants showed trajectories of hopeful future expectations that decreased across middle adolescence.

In the present study, we aimed to describe hopeful future expectations in both nomothetic and group differential terms. In other words, we sought to model average trajectories of hopeful future expectations across childhood through early adulthood in a large sample of participants from the United States; we also included predictors of slope and intercept in our model to examine whether certain demographic characteristics would predict the level and shape of these trajectories. There is no prior research, to our knowledge, of developmental differences in hopeful future expectations by race/ethnicity or gender (but see Seginer 2009, for information on the impact of gender and cultural socialization on the development of future orientation in youth).

The Current Study

There were two purposes of the present research. We sought to enhance knowledge about hopeful future expectations, a character virtue that is thought to be an important predictor of positive youth development (e.g., Schmid et al. 2011a, b; Callina et al. 2014b). However, the developmental literature supporting the theoretical role of hopeful future expectations in youth thriving has been limited because longitudinal data sets have, together, provided only unconnected assessments of its developmental course and contribution to thriving. Therefore, we had the second goal of employing IDA techniques to enhance understanding of this character virtue. It is our own hopeful future expectation that this innovation in developmental methodology will be a resource for other scholars seeking to widen the lens of the developmental course of character virtue development, which has previously only been examined in brief snapshots of development.

In addressing the two proposes of the research, we were fortunate to be able to capitalize on the presence of several longitudinal data sets available at the Institute for Applied Research in Youth Development. These data sets assess youth strengths within various in-school and out-of-school time settings that implement character education programs or related activities and curricula. The data sets derived from these studies include information about specific character virtues, such as gratitude, thrift, integrity, and hopeful future expectations; possible predictors of character development, such as program engagement and intentional self-regulation skills; and possible outcomes of character virtue development, such as youth contribution and civic engagement. The studies included participants ranging from age 6 through 30. We highlight those studies selected for inclusion in our analyses and their participants in greater detail below (see also Callina et al. 2016, for details about the theoretical foundations of the positive character development studies presented here).

In this article, we describe the IDA method and its application to four of the Institute for Applied Research in Youth Development projects most suitable for modeling character development across the first three decades of life. Generally, the purpose of IDA techniques is to combine data from independent existing studies and pool the data into a single data set for further analysis (see Curran and Hussong 2009; Hofer and Piccinin 2009; Shrout 2009). The IDA method proceeds as follows: first, researchers determine the "ultimate goal" of their analysis (Curran et al. 2008, p. 366). Next, they determine which measures to model by identifying overlapping items across the different studies (called "anchor" items; Curran and Hussong 2009, p. 19). Then, researchers evaluate heterogeneity due to measurement and sampling. Finally, the researchers conduct their analyses.

We describe the process of conducting IDA in greater detail, below. Issues such as the benefits and limitations of IDA will be discussed. The key concern in fitting a model across aggregated data sets within the IDA framework is careful consideration of between-study heterogeneity (Curran and Hussong 2009). Accordingly, we address heterogeneity due to sampling, history (i.e., age, period, cohort; Schaie 1965), and measurement. This latter issue—measurement—is our primary area of focus. Using examples from four of the above-noted studies, we demonstrate how IDA can be used to link data across multiple, independent samples to model pathways of character development across longer time spans and a greater number of participants than is possible in any single data set.

Method

Because this article is intended as a demonstration of how integrative data analysis can be applied to the study of the character virtue of hopeful future expectations in a manner that will generate new knowledge about its developmental course across the adolescent period, we describe our method and analysis in considerable detail and present the results of each step of the method and analysis along the way (in contrast to a conventional article in which an entire Method section is presented first, followed by a complete Results section). We chose this presentation format to provide the best illustration of the process by which we applied IDA to study hopeful future expectations. However, this approach to describing our methodology should be useful to enhance testing of other character virtues that have comparable theory-data discrepancies (Card in press). The extant composition of the longitudinal knowledge base does not enable evidence to be marshaled to either verify the developmental course of character virtues prior to, during, or after adolescence or to test ideas about the importance of these trajectories for positive development. In the present study, our ultimate aim was to use IDA methods to assess growth models in order to investigate individual variability in trajectories of hopeful future expectations and of other character virtues from late childhood to young adulthood. This process will differ, at least partially, for other constructs, other data sets, and other ultimate analysis aims. Interested readers should refer to Curran and Hussong (2009; see, too, Curran et al. 2008) for further details of conducting an IDA.

Each subsection below corresponds to a step in the process of conducting an IDA. Here, we describe our method for conducting each step, and then detail its results, before moving on to the next step.

Specifying a Research Question

IDA begins with the identification of an overarching research question or ultimate analysis aim. These ideas and questions guide the initial selection of the larger pool of studies from which the eventual final group of studies is chosen. For example, for our analyses, we began with the intention of investigating trajectories of character attributes from childhood to adolescence. These analyses are therefore exploratory; nevertheless, describing the course of development for specific measures of character attributes may provide insight to researchers seeking to use those measures in future character program evaluations.

Choosing Studies

With a guiding research question, researchers move to the identification of potentially appropriate data sets from which to pool participants into a single data set (Curran and Hussong 2009). The group of data sets may be large at this point and should include any available studies (similar to the process of conducting a meta-analysis; Card 2011).

We began by examining the codebooks of studies that have been conducted at the Institute for Applied Research in Youth Development. These studies share a focus on youth strengths and the assets in contexts that promote positive youth development and character virtues such as humility and integrity. In particular, projects at Institute for Applied Research in Youth Development from the past decade have included collaborations with Boy Scouts of America (Hilliard et al. 2014; Wang et al. 2015); 4-H (Lerner et al. 2005; Geldhof et al. 2013); a trade school for young men (Johnson et al. 2014); and the United States Military Academy at West Point (Callina et al. 2017). In addition, we have conducted research on the Arthur Interactive Media program in elementary schools (Bowers et al. 2015); sports participation in high school (Ferris et al. 2016); and the development of entrepreneurship among college and university students (e.g., Geldhof et al. 2014). Across the available studies, we selected for further consideration those investigations for which data collection has been completed (which eliminated several studies currently in progress).

Based on the above criteria, we selected four studies conducted through the Institute for Applied Research in Youth Development. These are (in order of average age of participants, from youngest to oldest): the Boy Scouts of America Character Assessment and Merit Project (CAMP); the 4-H Study of Positive Youth Development; the Assessment of Character in the Trades (ACT) study; and the Young Entrepreneurs Study (YES). Here we describe briefly the method by which data were collected for each of these studies, and refer the reader to other sources for more detailed information about each study. Figure 1 provides an overview of the age of participants and years of data collection for each study, and Fig. 2 illustrates the number of observations with hopeful future expectations data at each age included in the analyses, according to study group.

Character and merit project (CAMP)

CAMP was conducted within the greater Philadelphia area. The Boy Scouts of America Council within this region is the Cradle of Liberty Council. The Cradle of Liberty initiates its programs following the national mission of Boy Scouts of America programs throughout the U.S. The Cradle of Liberty serves 10,000 Scouts and is facilitated by volunteer leaders (mostly parents) from 250 packs throughout the region. In addition, the Cradle of Liberty includes various professional staff members who provide support to volunteer leaders throughout the region. Full details of the methodology of CAMP have been

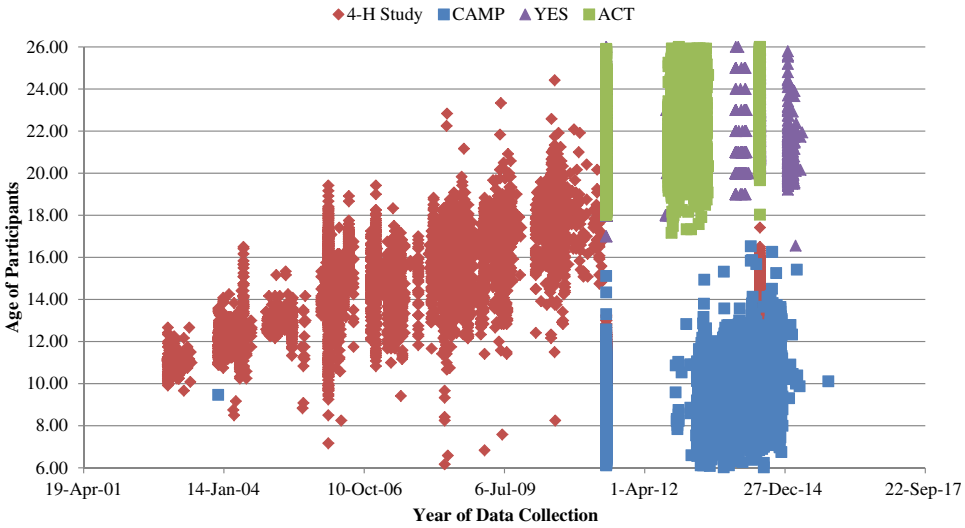

Fig. 1 Age of participants in each of the four studies and the years in which the data were collected. *CAMP* Character and Merit Project, *ACT* Assessment of Character in the Trades, *YES* Young Entrepreneurs Study

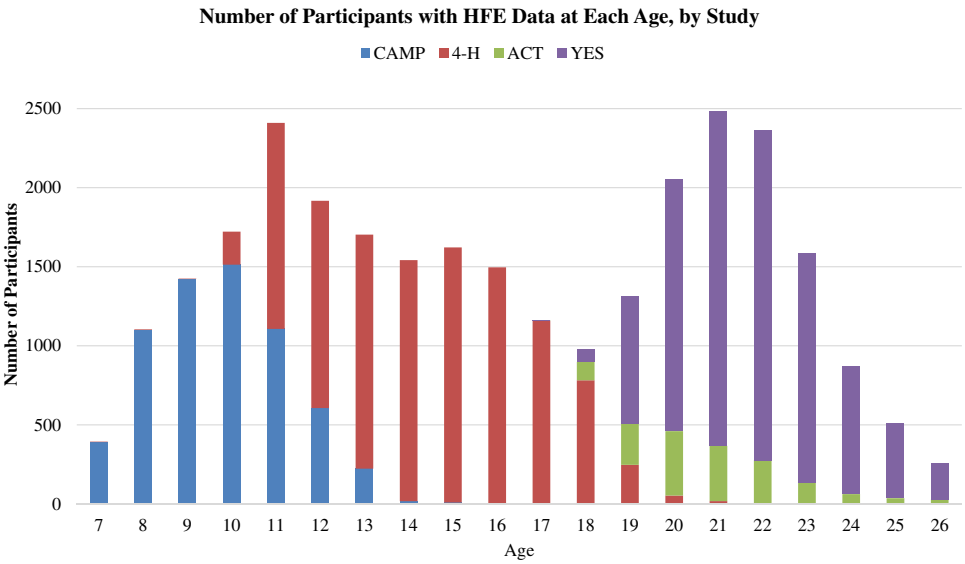

Fig. 2 Number of observations with Hopeful Future Expectations data at each age, according to study group. *CAMP* Character and Merit Project, *ACT* Assessment of Character in the Trades, *YES* Young Entrepreneurs Study

presented in prior reports (e.g., Hilliard et al. 2014; Wang et al. 2015).

CAMP procedure

Participant recruitment began in 2012 and ended in 2015, and included five waves of data collection, collected at approximately 6-month intervals. All youth were recruited from the same region in the greater Philadelphia metropolitan area. Scouts were recruited from regularly-scheduled Scouting meetings held in local community locations, including churches, community centers, or schools; the research team asked adult leaders of Cub Scout packs to help make parents aware of the study, collect parental consent, and administer the questionnaires during pack meetings. A comparison sample of youth was also recruited from public, charter, and/or Catholic schools in the greater Philadelphia area. Questionnaire materials were administered by pack leaders (in Cub Scout packs) or by Institute for Applied Research in Youth Development researchers in classrooms (for the comparison sample). Overall, 2751 youth participated and most participants were able to complete the survey within 15 min. Parents provided information about demographic characteristics, including participants' race/ethnicity. Youth received small toys and trinkets for their participation.

CAMP participants

The analytic sample for this study included a total of 2751 youth. Of these, about two-thirds were Boy Scouts and the other third were non-Scout participants (Wave 1: *M*age = 8.48, SD = 1.59; 80% male). Among participants for whom race/ethnicity information was available (approximately

33% of the total sample), 70% were White or European American, 10.2% were Black or African American, 6.2% were Hispanic or Latino, 3.1% were Asian or Pacific Islander, 1.8% self-rated as "Other", 8.2% identified as Multiethnic or Multiracial, and 0.6% were American Indian.

4-H Study of Positive Youth Development

The 4-H Study of Positive Youth Development is a longitudinal study of youth that began in 2002 and ended in 2012. Participants were surveyed about once each year from 5th Grade through 12th Grade. Full details of the methodology of the 4-H Study of Positive Youth Development can be found in prior publications (e.g., Lerner et al. 2005; Lerner et al. 2015).

4-H study procedure

For the first three waves of data collection, teachers or program staff obtained parental consent and the surveys were administered to youth by trained study staff or research assistants hired at more distant locations. A detailed protocol was used to ensure that data collection was administered uniformly and to ensure the return of all study materials. Beginning in Wave 2, students who did not complete the survey in person could receive a survey in the mail to complete and return. Beginning in Wave 5, youth were invited to complete the survey online, and online survey administration was the predominant method of data collection in Waves 6 through 8. Participants received gift cards for their participation.

4-H study participants

Overall, across all eight waves of the study, 7071 youth (59.9% female; Wave 1: *M*age = 10.97, SD = .53) in 42 states were surveyed. Participants who self-identified race were 68% White or European American, 9.7%; Hispanic or Latino, 7.5% Black or African American, 1.9% Asian, Asian American or Pacific Islander and 1.6% Native American or American Indian. An additional 2.4% self-identified as Multiethnic or Multiracial, 1.8% self-reported "Other" and 7.2% had selected an inconsistent race or ethnicity across the eight waves of the study.

Assessment of character in the trades (ACT) study

The ACT Study was a 3-year, mixed-method longitudinal study assessing character, citizenship, and vocational development in a sample of young men attending trade schools and community colleges in the greater Philadelphia area. Data collection began in August, 2012 and ended in April, 2015. Participants completed surveys about individual and contextual features of development; data were collected at the beginning, middle, and end of the 3-year period. Here, we focus on the quantitative or survey components of the data collection. Full details about the method may be found in Johnson et al. (2014).

ACT procedure

Participants were recruited from four post-secondary schools in the greater Philadelphia area with the support of the administration from each school. Surveys were administered electronically; about half the sample completed the survey during designated time in a computer lab, and the other half completed the survey at their own convenience. Researchers requested follow-up participation from participants via phone calls, postcards, and emails, and participants received gift cards to local retail stores for their involvement.

ACT participants

The sample consisted of 1328 participants (Wave 1: *M*age = 19.22, SD = 1.71). Participants who self-identified race were 74.6% White or European American, 8.5% Black or African American, 4.9% Asian or Pacific Islander, 2.9% Hispanic or Latino, and 0.1% Native American or American Indian. An additional 4.1% self-identified as Multiethnic or Multiracial, and 0.9% self-identified as "Other".

Young entrepreneurs study (YES)

The YES project was a 3-year, mixed-method longitudinal study aimed at understanding the development of entrepreneurship in a sample of post-secondary school students from colleges and universities located in New England, West Coast, and Mid-west regions of the United States. Full details about the methodology may be found in Geldhof et al. (2014).

YES procedure

Data were collected using a cohort-sequential design. At Wave 1, researchers recruited participants by contacting professors, administrators, and student organization leaders at approximately 50 universities and asked them to forward their students a recruitment email that contained a link to the YES survey. Participants either received course credit or were entered into a raffle for an iPad as compensation for their involvement in Wave 1. In Wave 2, participants had the option of being entered into a raffle for an iPad or for a $500 Amazon gift card. In addition, researchers offered returning participants a $35 Amazon gift card. In Wave 3, participants received a $35 Amazon gift card.

Approximately 1 year after completing the initial survey, participants who had provided their contact information were recruited for the Wave 2 data collection. To account for attrition, additional participants at Wave 2 were recruited using the same recruitment methods previously described. At Wave 3, participants who had completed surveys and provided contact information at either of the previous waves were re-contacted.

YES participants

The sample consisted of 9586 participants (53% Female; Wave 1: *M*age = 21.14, SD = 1.64). Participants who self-identified race were 54.2% White or European American, 21.1% Asian or Pacific Islander, 4.7% Black or African American, 5.9% Hispanic or Latino, and 0.1% Native American or American Indian. Those who self-identified as Multiethnic or Multiracial were 4.5 and 1.1% classified themselves as "Other."

Choosing Items

Once we identified which studies to include in the IDA, the next step was to choose the specific items to be used for analysis. The choice of items is guided by several different concerns. Particularly relevant to the present demonstration is the issue of heterogeneity due to measurement. According to Curran and Hussong (2009), "In most IDA applications, a key goal is to optimally capture the measurement of specific theoretical constructs both within each sample individually and, more importantly, within the aggregated sample as a whole" (p. 12). Because the focus of our analysis was on modeling growth curves of a particular character virtue, measurement invariance is especially important, and we will discuss our procedure for addressing invariance in greater detail, below.[1]

After selecting the studies, we used the codebook from each study to tabulate the various character virtues assessed and items included for each of the virtues. We evaluated which virtues were present in at least three of the studies. This process resulted in the identification of 12 attributes across the four data sets (i.e., CAMP, 4-H, YES, and ACT). We then further examined these attributes by comparing the items used to assess them in each study, with the aim of selecting an attribute that had at least a few overlapping items among studies, and which was known through prior analyses to have good psychometric properties. In IDA, it is necessary to identify items that overlap across studies—called "anchor" items (Curran and Hussong 2009, p. 19)—for the measurement invariance testing process.

Through this process, we identified one construct for further analysis: hopeful future expectations. Items related to hopeful future expectations were originally drawn from a set of questions within the 4-H Study data set that assessed participants' expectations that they will experience specific positive situations later in life (Schmid et al. 2011a, b). Participants were asked the following question: "Think about how you see your future. What are your chances for the following?" Items include being respected in the community, being healthy, having a job that pays well, and having a happy family life. In CAMP, the prompt was formatted differently for the younger participants: "Think about your future. What will your life be like when you grow up?"

There is a total of 15 hopeful future expectations items with a response format ranging from 1 = *very low* to 5 = *very high*. Higher scores indicate higher expectations of the likelihood that certain future outcomes will occur. The results of prior research (e.g., Schmid et al. 2011a) have indicated that this measure has good psychometric properties; Cronbach's alpha ranges from 0.71 to 0.95 across the four studies reported here. In addition, analyses from the 4-H Study find that hopeful future expectations has theoretically—expected within- and across-time relations with other variables, including positive associations with the "Five Cs" of Positive Youth Development, youth contribution, and intentional self-regulation skills, and negative associations with depressive symptoms and risk behaviors (Schmid et al. 2011a, b). Moreover, findings indicate that youth with high levels of hopeful future expectations are more effective in their ability to form positive relationships with important non-parental adults (e.g., Bowers et al. 2012; Bowers and Johnson 2013), which makes it a potentially key variable for understanding positive youth development.

We define character attributes as the set of behaviors, attitudes, and emotions that allow an individual to thrive in a particular context (Berkowitz 2011; Lapsley and Narvaez 2006; Lerner and Callina 2014; Sokol et al. 2010). As an individual asset contributing to the positive development of youth, hopeful future expectations are a representative character attribute. Hopeful future expectations are thought to work closely with goal-setting and management skills in providing the emotional and motivational energy needed to garner resources and engage in positive activities in the present, in order to realize positive goals for the future (Callina et al. 2014a; Schmid and Lopez 2011). Indeed, its robust association with other character attributes—combined with its good psychometric properties—is likely why the hopeful future expectations items were included in so many studies of character and positive development at the

[1] Curran and Hussong (2009) distinguish between measurement *invariance*, which applies to studies with the same items, and measurement *comparability*, which applies to studies with different but similar items. For this demonstration, we selected a construct that shares the same items across studies, and so we focus on measurement invariance here.

Institute for Applied Research in Youth Development. For these reasons, we believed hopeful future expectations to be a good candidate for this IDA demonstration.

Preparing the Data Set

Once the items of interest from each study were identified, the next step for IDA was to create the complete data set for the analyses. We tested measurement equivalence within and across studies, and conducted growth curve models to examine trajectories of hopeful future expectations. The process of preparing the data set will be different depending on the structure of the data sets involved and the eventual aim of the IDA that will be conducted. Here, we describe considerations we identified during this process, with the purpose of providing a pedagogical illustration. Some of the considerations are analytical (e.g., identification of moderators), whereas others are more practical.

For analytical issues, we first considered other variables—beyond the hopeful future expectations items—to include in the data set as potential moderators. One such variable should be a variable indicating study participation (here, for example: CAMP, 4-H Study, YES, ACT) to enable examination of between-study differences. Other moderators that are included will depend on the nature of the studies involved, and also on the nature of eventual analyses that will be conducted. For the present analyses, we included gender and race as moderators predicting intercept and slope of the growth curve models. Although there is no prior research on gender and race differences in trajectories of hopeful future expectations, per se, Seginer (2009) provides an excellent treatment on the impact of gender and cultural socialization on the development of future orientation in youth.

One challenge for creating a data set that would be appropriate for invariance testing (i.e., testing for measurement equivalence first across ages within study and then between studies where there was overlap in the ages of the samples) was structuring the data in a way that would allow us to examine the hopeful future expectations measure at different ages. The longitudinal data sets chosen for inclusion in our IDA were structured by wave of data collection. In many instances, this structure causes the x-axis to be divided by occasion of measurement, and does not indicate participants' ages or timing of the data collection. Although there are specific types of longitudinal data analyses in which structuring the data set by wave may be appropriate (Lerner et al. 2009), a wave-based structure was not appropriate in this case. If we retained the wave-based structure of our original data sets, many ages would be represented at each wave, which would preclude our ability to examine measurement equivalence across age groups. Therefore, for the invariance testing, we restructured the data sets so that each variable represented the repeated measurement of hopeful future expectations across age of participants (rounded to the nearest age in years), rather than wave of data collection. For example, in the CAMP study, we created variables for hopeful future expectations at age 7, age 8, and so on. However, in the second phase of the analysis, modeling the growth curve trajectories, we used the original wave-based structure of the data sets, because we used the tscores function in MPlus (Muthén and Muthén 1998–2010) to model individually varying times of observation.

The issue of age and cohort was another major consideration for preparing our data set. If the pooled data set contained a large number of birth cohorts, one option would have been to include cohort (in addition to age), and treat it as a between-participants variable in our analyses. However, the four studies used in the IDA for this demonstration were all conducted within a 13-year period, with only two distinguishable birth cohorts represented: participants born in the early 1990s (4-H Study, ACT, and YES) and participants born in the early 2000s (CAMP). Figure 1 illustrates the timing of data collection by age of study participation. Therefore, we decided that variables indicating participants' study and age would account for any variation due to cohort differences, and did not include cohort itself as a variable in these analyses.

Some other issues we identified during this process were more pragmatic. For example, the data sets had to be structured in a way that would allow us to merge them in SPSS, which involved renaming variables, setting the variable types and scales to be identical across the four data sets, and creating unique identifiers for each participant across the four studies. In addition, we identified demographic characteristics from each study that allowed us to describe the sample using more or less comparable features (participants are described in the Method section above). Once we had a final data set with data from all four studies, we turned to the analysis phase.

Analysis and Results

Selecting the appropriate analysis required answering two questions: What is the ultimate analysis goal? And, how should the researchers account for between-study differences? In response to the first question, the ultimate analysis goal of this IDA demonstration was to examine trajectories of hopeful future expectations, a character attribute that contributes to the positive development of youth and which may have empirical value for researchers studying character because of its good reliability and validity.

The second analysis question, about how researchers should account for between-study differences, is a bit more

complex, and we have already pointed to some of the practical issues of preparing the pooled data set, above. Because IDA, by design, includes data from several studies, researchers must decide how to investigate heterogeneity due to study membership. We chose to use a fixed-effects approach to IDA (Curran and Hussong 2009). In this approach, researchers use dummy codes to denote study membership as a fixed characteristic of each individual observation. We chose the CAMP study as the reference group for the dummy-coded study variable because it had the lowest average age. The advantage of this technique is that it partially addresses potentially unmeasured sources of between-study variation, such as heterogeneity due to sampling, history (e.g., cohort differences), geography, or other design characteristics that are unique to each study (including method of data collection, which varied across the four studies). The disadvantage to the fixed-effects approach of IDA—as compared to the random-effects, or multilevel, approach to IDA—is that it removes between-sample variability from the model. Nevertheless, we determined that fixed-effects IDA would be appropriate for the purposes of this demonstration.

Clearly, the analyses for IDA involve considerably more preparation than single-study analyses. In this demonstration, for instance, we determined that the data files would need to be "reshaped" so that the items were structured by age instead of wave. For example, in the original data sets variable "w1HFE04" indicated Item 4 for the hopeful future expectations scale at Wave 1 data collection. However, participants in each study were different ages when the data were collected (e.g., a participant in CAMP might have been 7 years old at Wave 1 and a participant in the 4-H Study would likely have been 10 years old at Wave 1). Therefore, we reshaped the data file so that each item indicated the hopeful future expectations score for each participant at each age, rounded to the nearest integer (e.g., "a7HFE04" indicated the Item 4 hopeful future expectations score at age 7).

It is important to note that the process of conducting the main analysis (in this case, growth curve modeling) with a pooled data set is exactly the same as if the analysis was being conducted using data from a single study, with the exception that between-study variables may be included as covariates in the analysis. The software program (in this case, data were analyzed in Mplus) does not "know" that the data are from an IDA data set!

The analysis itself involved two stages. First, we established measurement equivalence within each study across the longitudinal waves of data collected, then we established measurement equivalence across the four studies. Next, we examined the trajectories of hopeful future expectations. Data cleaning and manipulation were conducted in SPSS, and the analyses were conducted in Mplus (Muthén and Muthén 1998–2010). Syntax and data files are available from the authors upon request.

Establishing Measurement Equivalence

We evaluated how the hopeful future expectations construct had been operationalized across the four studies. This process is called checking for "heterogeneity due to measurement" (Curran et al. 2008, p. 369; see also Curran and Hussong 2009, p. 19). This process may be conducted iteratively with the process of selecting studies described above. In this demonstration, the goal of addressing heterogeneity due to measurement was to calculate scale scores for further analysis, and we proceeded by checking the following (these steps may differ slightly if measurement comparability is also necessary; see Curran and Hussong 2009): 1. Dimensionality, which involves conducting factor analyses; 2. Measurement equivalence; and 3. Scoring (i.e., calculating individual- and time-specific scale scores for every participant at every time point at which they were assessed). Tables 1 and 2 show the hopeful future expectations items that were available in each data set and the items that were selected for inclusion in each longitudinal invariance test.

Longitudinal invariance was established using the recommendations outlined by Little (2013), in which a longitudinal Confirmatory Factory Analysis is employed to test whether a construct and its items are comparable across time. For this study, rather than testing invariance simultaneously across all ages within a study, we separately tested invariance among smaller, overlapping age groupings (see Table 2). For example, we tested invariance of the hopeful future expectations items within CAMP study participants by ages 7–9; 9–11; and 11–13.

If longitudinal invariance is established, researchers can reasonably assume that respondents in each age group attribute the same meaning to the latent hopeful future expectations construct, and that the intercepts of the hopeful future expectations items are the same in each age group; thus, longitudinal comparisons of the latent hopeful future expectations construct are valid (van de Schoot et al. 2012). The longitudinal Confirmatory Factor Analysis method proceeds in four steps. First, we fit an alternative null model (Widaman and Thompson 2003) to use in further model testing, rather than the one provided by Mplus. The purpose of the null model is to generate a chi-square value by which to compare subsequent models. Our alternative null model, as described by Little (2013), included not only the assumption of no covariances among items (the standard null model) but also specified that the means and variances of the same items were equivalent over measurement occasions. This specification is most appropriate for longitudinal invariance testing.

Table 1 Hopeful future expectations items included in each data set at each wave of data collection

Variable name and stem		ACT			CAMP					YES			4-H Study							
		W1	W2	W3	W1	W2	W3	W4	W5	W1	W2	W3	W1	W2	W3	W4	W5	W6	W7	W8
HFE01	Be able to buy the things you need.	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓			
HFE02	Be able to do things you want.	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓			
HFE03	Have a job you like doing.	✓	✓	✓						✓			✓	✓	✓	✓	✓			
HFE04	Be healthy.	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
HFE05	Have a job that pays well.	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓			
HFE06	Be able to live wherever you want.	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓			
HFE07	Have a happy family life.	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓			
HFE08	Be safe.	✓		✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HFE09	Have friends you can count on.	✓		✓						✓			✓	✓	✓	✓	✓	✓	✓	✓
HFE10	Go for additional college.	✓		✓									✓	✓	✓	✓	✓			
HFE11	Be respected in the community.	✓		✓						✓			✓	✓	✓	✓	✓			
HFE12	People will think I am a good person.				✓	✓	✓	✓	✓											
HFE13	Graduate from high school.												✓	✓	✓	✓	✓			
HFE14	Go to college.												✓	✓	✓	✓	✓			
HFE15	Graduate from college.												✓	✓						

ACT Assessment of Character in the Trades, *CAMP* Character and Merit Project, *YES* Young Entrepreneurs Study

Second, configural invariance is established, which indicates whether the patterns of factor loadings (that is, the patterns of free and fixed parameters) are the same across time. Next, weak factorial invariance is assessed by testing whether the factor loadings themselves are the same across time. Finally, the intercepts of the factor loadings are constrained to be equal to test for strong factorial invariance across time. Invariance is established if the model fit does not change substantially in each subsequent invariance test, using the criteria that a change of less than 0.01 in the Comparative Fit Index provides the best support for invariance (Cheung and Rensvold 2002).

Measurement invariance was established across different age groups within each study data set first, and then we tested for invariance within the pooled data set. In this way, we were able to establish measurement equivalence for hopeful future expectations by both age and study. Table 2 shows the items that were used in each between-study invariance test. We tested invariance of the hopeful future expectations items at age 11 between 4-H and CAMP participants; at age 18 between 4-H and ACT participants; and at age 20 between ACT and YES participants. These age groups were chosen to maximize coverage (i.e., the number of participants with data at both time points included in the analyses). Tables 3a–d show the results of these analyses.

The between-study invariance proceeded using the same steps as the longitudinal invariance, with the exception of fitting the alternative null model. In most cases (with an exception described below) we used the default null model provided by Mplus because it is reasonable for between-group invariance testing. For all between-study invariance tests, we used Full Information Maximum Likelihood estimation.

In two cases, the between-study invariance tests proceeded slightly differently. In the case of the CAMP and 4-H invariance tests, only two hopeful future expectations items were present in both data sets (items HFE04 and HFE07); in the case of the 4-H and ACT invariance tests, only three items overlapped (items HFE04, HFE08, and HFE09; see Table 2). When models include only a single factor within each group (i.e., hopeful future expectations), model identification requires at least three indicators, and four indicators are needed to produce model fit statistics that are useful for comparison (three indicators produces a just-identified model). Because these between-study instances of invariance testing included only two overlapping indicators, we chose to generate random data for the other indicators using a procedure described by Geldhof et al. (2013) and Widaman et al. (2013). The factor loadings of these randomly generated indicators were fixed at zero and their means were freely estimated in our models. As described by Geldhof et al. (2013), randomly generated indicators necessarily cause small amounts of misfit because they do not correlate with all other indicators perfectly at zero. Therefore, we estimated residual covariances between the randomly generated indicators and all real indicators in the

Table 2 Hopeful future expectations items used in the longitudinal and between-study measurement invariance test

Study	Age Group	HFE01	HFE02	HFE03	HFE04	HFE05	HFE06	HFE07	HFE08	HFE09	HFE10	HFE11	HFE12
Longitudinal invariance tests													
CAMP	7–9				✓✓			✓✓					✓✓
	9–11				✓✓			✓✓					✓✓
	11–13				✓✓			✓✓					✓✓
4-H	10–12	✓	✓		✓✓	✓	✓	✓	✓✓	✓✓			
	12–14	✓	✓		✓✓	✓	✓	✓	✓✓	✓✓			
	14–16	✓	✓		✓✓	✓	✓	✓	✓✓	✓✓			
	16–18				✓✓				✓✓	✓✓			
	18–20				✓✓				✓✓	✓✓			
ACT	18 and 20	✓✓	✓✓		✓✓	✓		✓✓	x	x			
	20–22	✓✓	✓✓		✓✓	✓	✓	✓✓					
	22–23	✓✓	✓✓		✓✓	✓	✓	✓✓					
YES	18–19	✓✓	✓✓			✓	✓✓		✓✓				
	19–21	✓✓	✓✓			✓	✓✓		✓✓				
	21–23	✓✓	✓✓			✓	✓✓		✓✓				
	23–25	✓✓	✓✓			✓	✓✓		✓✓				
	25–26	✓✓	✓✓			✓	✓✓		✓✓				
Between-study invariance tests													
CAMP & 4-H (age 11)		✓	✓		#	✓	✓	#	✓	✓			✓
4-H & ACT (age 18)		✓	✓		#	✓	*	✓	#	#			
ACT & YES (age 20)		*	*		#	#	*	#	#	*			

CAMP Character and Merit Project, *ACT* Assessment of Character in the Trades, *YES* Young Entrepreneurs Study

An "x" indicates that the item was available in the data set but was not included in the invariance testing due to poor covariance coverage

Check marks (✓) indicate available items. Double check marks (✓✓) indicate items that were used to compute scale scores in the present study

Asterisks (*) indicate that items were removed because modification indices indicated that the item was highly correlated with other items

Pound sign (#) indicates, for between-study invariance tests, overlapping items that were directly tested for invariance

alternative null model for these tests, and in subsequent steps of invariance testing. The means of the randomly generated indicators were also freely estimated within each group.

Table 4 shows the model fit statistics for the configural model, factor loading (weak) invariant model, and intercept (strong) invariant model for hopeful future expectations in the between-study tests. Results indicated that the latent hopeful future expectations construct showed configural, weak, and strong invariance, which means that the pattern and magnitude of the factor loadings for each factor, and the item intercepts, were the same across age groups, with one exception. The ACT/YES invariance comparisons indicated that intercept invariance was not reasonable for these data, given the large change in Comparative Fit Index when that constraint was imposed. We inspected the results of the loading invariance model and identified item HFE05 as a potentially non-invariant item. We then specified a partial intercept model (in which the intercepts of HFE05 were not constrained across the two groups), and this model passed the intercept invariance test. Given that result, we did not include item HFE05 in our computation of scale scores (described below).

Model fit statistics suggested acceptable fit for most models in both the longitudinal (Tables 3a–d) and between-study invariance tests (see Table 4). In some cases, model fit was improved by removing an item with poor covariance coverage or if modification indices indicated that the item was highly correlated with other items. For example, in the ACT Study, item HFE08 ("Be safe") was missing for many participants, and item HFE06 ("Be able to live wherever you want") was strongly correlated with several other items. For the strong test of invariance between ACT and YES, however, the model fit was mediocre (RMSEA = .099). Modification indices indicated that correlating the residual variances of all of the hopeful future expectations items in the YES data set would have improved the model fit. However, because the ultimate aim of our analyses was to compute scale scores for modeling growth curves, we opted to leave the residual variances uncorrelated.

To compute the hopeful future expectations scale scores, we simply used the average of hopeful future expectations

Table 3a Model fit from longitudinal configural, loading, and intercept measurement invariance models for the hopeful future expectations items in CAMP

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI	Pass? (ΔCFI ≤ .01)
Ages 7 to 9, $N = 1591$						
0. Null	1360.811 (48)	<.001	0.131 (.125 to .137)	.000	.232	
1. Configural	33.611 (15)	<.01	0.028 (.015 to .041)	.986	.965	
2. Loading	39.251 (19)	<.01	0.026 (.014 to .037)	.985	.970	PASS (ΔCFI = .001)
3. Intercept	44.229 (23)	<.01	0.024 (.013 to .035)	.984	.974	PASS (ΔCFI = .001)
Ages 9 to 11, $N = 1879$						
0. Null	1967.141 (48)	<.001	0.146 (.140 to .151)	.000	.244	
1. Configural	17.828 (15)	.2718	0.010 (.000 to .025)	.999	.996	
2. Loading	25.498 (21)	.0591	0.017 (.000 to .028)	.997	.990	PASS (ΔCFI = .002)
3. Intercept	33.794 (23)	.0682	0.016 (.000 to 0.27)	.995	.991	PASS (ΔCFI = .002)
Ages 11 to 13, $N = 942$						
0. Null	887.194 (48)	<.001	0.136 (.128 to .144)	.000	.206	
1. Configural	36.427 (15)	<.01	0.039 (.023 to .055)	.973	.935	
2. Loading	51.224 (19)	<.001	0.039 (.024 to .053)	.968	.936	PASS (ΔCFI = .006)
3. Intercept	46.923 (23)	<.01	0.033 (.019 to .047)	.970	.953	PASS (ΔCFI = −.004)

RMSEA Root Mean Square Error of Approximation, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index

Table 3b Model fit from longitudinal configural, loading, and intercept measurement invariance models for the hopeful future expectations items in the 4-H study

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI	Pass? (ΔCFI ≤ .01)
Ages 10 to 12, $N = 2186$						
0. Null	11594.592 (308)	<.001	0.129 (.127 to .131)	.000	.096	
1. Configural	1279.800 (225)	<.001	0.046 (.044 to .049)	.906	.884	
2. Loading	1333.771 (239)	<.001	0.046 (.043 to .048)	.903	.885	PASS (ΔCFI = .004)
3. Intercept	1406.234 (253)	<.001	0.046 (.043 to .048)	.898	.888	PASS (ΔCFI = .005)
Ages 12 to 14, $N = 2788$						
0. Null	21759.693 (308)	<.001	0.158 (.156 to .160)	.000	.101	
1. Configural	1840.236 (225)	<.001	0.051 (.049 to .053)	.924	.907	
2. Loading	1855.479 (239)	<.001	0.049 (.047 to .051)	.924	.913	PASS (ΔCFI = .000)
3. Intercept	1902.867 (253)	<.001	0.048 (.046 to .050)	.923	.916	PASS (ΔCFI = .001)
Ages 14 to 16, $N = 3512$						
0. Null	16834.634 (238)	<.001	0.141 (.139 to .143)	.000	.116	
1. Configural	1612.988 (225)	<.001	0.042 (.040 to .044)	.917	.915	
2. Loading	1632.096 (239)	<.001	0.041 (.039 to .043)	.916	.920	PASS (ΔCFI = .000)
3. Intercept	1647.423 (253)	<.001	0.040 (.038 to .041)	.916	.924	PASS (ΔCFI = .000)
Ages 16 to 18, $N = 2720$						
0. Null	4894.521 (48)	<.001	0.193 (.188 to .197)	.000	.250	
1. Configural	23.685 (15)	.0706	0.015 (.000 to .025)	.998	.996	
2. Loading	26.945 (19)	.1059	0.012 (.000 to .022)	.998	.998	PASS (ΔCFI = .000)
3. Intercept	32.001 (23)	.1001	0.012 (.000 to .021)	.998	.997	PASS (ΔCFI = .000)
Ages 18 to 20, $N = 905$						
0. Null	1653.446 (48)	<.001	0.192 (.184 to .200)	.000	.235	
1. Configural	35.871 (15)	<.01	0.039 (.023 to .056)	.987	.968	
2. Loading	41.992 (19)	<.01	0.037 (.022 to .052)	.986	.967	PASS (ΔCFI = .001)
3. Intercept	54.168 (25)	<.001	0.036 (.023 to .049)	.982	.973	PASS (ΔCFI = .004)

RMSEA Root Mean Square Error of Approximation, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index

Table 3c Model fit from longitudinal configural, loading, and intercept measurement invariance models for the hopeful future expectations items in ACT

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI	Pass? (ΔCFI ≤ .01)
Ages 18 and 20, $N = 446$						
0. Null	1317.252 (55)	<.001	0.227 (0.216 to 0.238)	.000	.179	
1. Configural	53.967 (29)	<.01	0.044 (0.025 to 0.062)	.980	.969	
2. Loading	58.811 (33)	<.01	0.042 (0.024 to 0.059)	.980	.973	PASS (ΔCFI = .000)
3. Intercept	61.986 (37)	<.01	0.039 (0.021 to 0.055)	.980	.976	PASS (ΔCFI = .000)
Ages 20 to 22, $N = 779$						
0. Null	2921.640 (177)	<.001	0.141 (0.137 to 0.146)	.002	.137	
1. Configural	250.992 (114)	<.001	0.039 (0.033 to 0.046)	.950	.933	
2. Loading	256.452 (124)	<.001	0.037 (0.031 to 0.043)	.952	.939	PASS (ΔCFI = .002)
3. Intercept	265.163 (134)	<.001	0.035 (0.029 to 0.042)	.952	.946	PASS (ΔCFI = .000)
Ages 22 and 23, $N = 310$						
0. Null	1115.871 (78)	<.001	0.207 (0.196 to 0.218)	.000	.149	
1. Configural	124.996 (47)	<.001	0.073 (0.058 to 0.089)	.924	.894	
2. Loading	130.271 (52)	<.001	0.070 (0.055 to 0.085)	.924	.904	PASS (ΔCFI = .000)
3. Intercept	132.973 (57)	<.001	0.066 (0.051 to 0.080)	.926	.915	PASS (ΔCFI = .002)

RMSEA Root Mean Square Error of Approximation, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index

Table 3d Model fit from longitudinal configural, loading, and intercept measurement invariance models for the hopeful future expectations items in YES

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI/ NNFI	Pass? (ΔCFI ≤ .01)
Ages 18 and 19, $N = 811$						
0. Null	1264.034 (55)	<.001	0.165 (0.157 to 0.173)	.000	.180	
1. Configural	80.349 (29)	<.001	0.047 (0.035 to 0.059)	.958	.934	
2. Loading	86.889 (33)	<.001	0.045 (0.034 to 0.056)	.956	.942	PASS (ΔCFI = .002)
3. Intercept	90.550 (37)	<.001	0.042 (0.031 to 0.053)	.956	.946	PASS (ΔCFI = .000)
Ages 19 to 21, $N = 3092$						
0. Null	6867.491 (125)	<.001	0.132 (0.129 to 0.135)	.000	.156	
1. Configural	278.775 (72)	<.001	0.030 (0.027 to 0.034)	.969	.955	
2. Loading	284.524 (80)	<.001	0.029 (0.025 to 0.032)	.969	.960	PASS (ΔCFI = .000)
3. Intercept	303.528 (88)	<.001	0.028 (0.025 to 0.032)	.968	.962	PASS (ΔCFI = .001)
Ages 21 to 23, $N = 3666$						
0. Null	7984.018 (125)	<.001	0.131 (0.129 to 0.133)	.000	.159	
1. Configural	354.852 (72)	<.001	0.033 (0.029 to 0.036)	.964	.947	
2. Loading	361.731 (80)	<.001	0.031 (0.028 to 0.034)	.964	.953	PASS (ΔCFI = .000)
3. Intercept	370.751 (88)	<.001	0.030 (0.027 to 0.033)	.964	.957	PASS (ΔCFI = .000)
Ages 23 to 25, $N = 1827$						
0. Null	3783.908 (120)	<.001	0.129 (0.126 to 0.133)	.000	.123	
1. Configural	269.702 (72)	<.001	0.039 (0.034 to 0.044)	.946	.921	
2. Loading	277.323 (80)	<.001	0.037 (0.032 to 0.042)	.946	.931	PASS (ΔCFI = .000)
3. Intercept	281.944 (88)	<.001	0.035 (0.030 to 0.039)	.947	.937	PASS (ΔCFI = −.001)
Ages 25 and 26, $N = 496$						
0. Null	802.311 (36)	<.001	0.207 (0.195 to 0.220)	.006	.227	
1. Configural	29.713 (15)	<.05	0.044 (0.020 to 0.068)	.981	.964	
2. Loading	31.124 (18)	<.05	0.038 (0.013 to 0.060)	.983	.977	PASS (ΔCFI = −.003)
3. Intercept	32.874 (21)	<.05	0.034 (0.004 to 0.055)	.985	.979	PASS (ΔCFI = −.001)

RMSEA Root Mean Square Error of Approximation, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index

Table 4 Model fit from configural, loading, and intercept measurement invariance models for between-study invariance tests

Model	Chi-square (df)	*p*	RMSEA [90% CI]	CFI	TLI	Pass? (ΔCFI ≤ .01)
4-H and CAMP, $N = 9822$						
1. Configural	367.717 (35)	<.001	0.044 (0.040 to 0.048)	0.926	0.848	
2. Loading	369.147 (36)	<.001	0.043 (0.039 to 0.047)	0.926	0.852	Pass (ΔCFI = .000)
3. Intercept	394.229 (37)	<.001	0.044 (0.040 to 0.048)	0.921	0.845	Pass (ΔCFI = .005)
4. Exchangeability	504.540 (45)	<.001	0.046 (0.042 to 0.049)	0.898	0.836	Fail (ΔCFI = .023)
4-H and ACT, $N = 7187$						
1. Configural	68.904 (20)	<.001	0.026 (0.020 to 0.033)	0.970	0.936	
2. Loading	70.343 (22)	<.001	0.025 (0.018 to 0.031)	0.970	0.943	Pass (ΔCFI = .000)
3. Intercept	72.384 (24)	<.001	0.024 (0.018 to 0.030)	0.970	0.947	Pass (ΔCFI = .000)
4. Exchangeability	89.542 (30)	<.001	0.024 (0.018 to 0.029)	0.963	0.948	Pass (ΔCFI = .007)
ACT and YES, $N = 1809$						
1. Configural	27.378 (4)	<.001	0.080 (0.054 to 0.110)	0.975	0.924	
2. Loading	28.100 (7)	<.001	0.058 (0.036 to 0.081)	0.977	0.961	Pass (ΔCFI = .000)
3. Intercept	98.611 (10)	<.001	0.099 (0.082 to 0.117)	0.904	0.885	Fail (ΔCFI = .071)
3. Partial Intercept	31. 568 (9)	<.001	0.053 (0.033 to 0.073)	0.976	0.968	Pass (ΔCFI = .001)
4. Exchangeability	34.322 (12)	<.001	0.045 (0.028 to 0.064)	0.976	0.976	Pass (ΔCFI = .000)

RMSEA Root Mean Square Error of Approximation, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index

items that met longitudinal and between-study invariance for each set of studies. These items were: HFE04, HFE07, and HFE12 (CAMP); HFE04, HFE08, and HFE09 (4-H Study); HFE01, HFE02, HFE04, and HFE07 (ACT); and HFE01, HFE02, HFE06, and HFE08 (YES).[2]

Examining Trajectories of Hopeful Future Expectations

To examine trajectories, we fit growth curve models. Growth curve models are a statistical technique for exploring within-person change and between-person differences in change (Grimm and Ram 2012; McArdle 2009; McArdle and Nesselroade 2003). Specifically, the model estimates the best-fitting overall pattern of change, indicated by the intercept (i.e., starting point) and slope (i.e., rate of change; Grimm and Ram 2012). Growth curve models commonly include estimates of a linear slope, and may also include a quadratic slope (a parameter which indicates how much the linear rate of change is, itself, changing). We specified both a linear and a quadratic slope in our models.

In addition to point estimates for the intercept and slope, growth curve models can also include variation around these estimates. We initially specified variation in the intercept and linear slope, but the models with variation in the linear slope produced estimation errors. Inspection of the results showed that there was no estimated variance in the linear slope, so we constrained the variance to be zero, which indicates that the model-implied estimate of the linear slope is the same for all participants. We retained the variance in the intercept.

In order to model the hopeful future expectations scores by age (rather than by wave of data collection), we changed the metric of the x-axis by using the tscores command in the Mplus, which allows the researcher to run a growth model with individually varying times of observation (if the data are in a wide format; Muthén and Muthén 1998–2010). The tscores option, however, does not produce estimates of model fit (Grimm and Ram 2012).

We also scaled the age variable such that zero represented the youngest age present in the data set (7 years) by subtracting 7 from each age variable. When age is scaled in this way, the resulting estimates are more interpretable and developmentally meaningful; here, the intercept represented the average model-implied score of hopeful future expectations when participants were 7 years old.

Results of the growth curve model are shown in Table 5. The average intercept for hopeful future expectations scores was 4.681 ($p < .001$), representing the expected hopeful future expectations score when participants were 7 years old, with significant variance around this estimate (0.098, $p < .001$). This intercept score is quite high (recall that hopeful future expectations is scored on a 5-point likert-type scale in these studies), although not surprising given

[2] Scale averages assume all items are exchangeable, meaning that they have equal factor loadings (DeShon, 2004). Although this assumption is usually not tested, we did examine whether it was feasible in our studies, to provide additional justification for the computation of scale scores. For this test, we began with the models assuming strong invariance and added the additional constraint of equality of factor loadings. The exchangeability constraint produced negligible decreases in model fit (based on the change in the Comparative Fit Index, ΔCFI ≤ .001) for the ACT/4-H and ACT/YES comparisons, but not for the CAMP/4-H comparisons (ΔCFI = .023); see Table 4.

Table 5 Results from growth curve analyses using tscores

	Estimate	SE	*p*-value
Model 1 ($N = 12,135$)			
Means			
Intercept	4.681	0.018	<.001
Linear Slope	−0.047	0.005	<.001
Quadratic Slope	0.000	0.000	.794
Variances			
Intercept	0.098	0.007	<.001
Linear Slope		Fixed at Zero	
Quadratic Slope		Fixed at Zero	
Model 2 ($N = 11,129$)			
Means			
Intercept	4.663	0.023	<.001
Linear Slope	−0.058	0.007	<.001
Quadratic Slope	0.001	0.000	.004
Variances/Residual Variances			
Intercept	0.091	0.007	<.001
Linear Slope		Fixed at Zero	
Quadratic Slope		Fixed at Zero	
Predicting Intercept			
Female	0.215	0.014	<.001
Non-White	0.025	0.015	.094
Study (CAMP is the Reference Category)			
4-H	−0.103	0.021	<.001
ACT	0.118	0.043	.005
YES	−0.329	0.039	<.001

CAMP Character and Merit Project, *ACT* Assessment of Character in the Trades, *YES* Young Entrepreneurs Study

previous findings that showed CAMP study participants' scores on hopeful future expectations were about 4.5 at Wave 1 of that study (Wang et al. 2015). The average linear slope was negative (−0.047) and significantly different from zero ($p < .001$), indicating that hopeful future expectations scores decreased on average across the ages included in the model. Again, this finding is consistent with prior research (Wang et al. 2015; Callina et al. 2014a). Although this average slope value was small, over the 19-year period included in the model it represents a decrease of nearly a point (−0.893). The quadratic slope was not significantly different from zero in this model.

In the second growth curve model we tested, we included three predictors of the intercept (we did not include predictors of the slopes because they did not have a statistically significant amount of variance). The three predictors were: sex (female vs. male), race (white vs. non-white), and study. We chose to dichotomize race in the analysis due in part to the fact that the percentage of any single non-white participant group was very small in some of the study samples. The study variable was represented by three dummy codes, with CAMP as the reference group.

In the second model, the estimates of the intercept (4.663, $p < .001$) and linear slope (−0.058, $p < .001$) were similar to the model with no predictors. The quadratic slope, in contrast, was statistically significant (0.001, $p = .004$). In this model, the intercept represents the predicted hopeful future expectations score for a 7-year old participant, who was male, White, and a participant in the CAMP study. This score would be expected to decrease by an average of −0.058 points per year, although the quadratic slope indicates that this decrease became steeper over time.

Sex was a significant predictor of the intercept ($b = .215$, $p < .000$), indicating that girls had a higher average score at 7 years old than boys. Identifying as White vs. non-White did not predict the intercept. We also observed study-based differences in the intercept, with ACT participants having higher average scores compared to CAMP participants, and 4-H and YES participants having lower average scores compared to CAMP participants.

We believe it is important to note here that the analytic sample for this second growth curve model was large ($N = 11,129$), but it was just more than half of the total number of participants in the pooled data set ($N = 20,736$). Cases were excluded from the growth curve model analysis if they were missing on all hopeful future expectations variables ($N = 2354$); missing on time scores (i.e., age; $N = 3746$) or missing on the predictors ($N = 3507$). We discuss the missing data in greater detail, below.

Discussion

In the present study, we sought to advance the knowledge base regarding the development of hopeful future expectations using an innovative methodological approach. This approach, called *integrative data analysis* (IDA), allows researchers to pool raw data across different studies in order to fit one model to an aggregated data set. Data for this article were derived from four longitudinal studies conducted by the Institute for Applied Research in Youth Development. This work demonstrates that extant longitudinal data sets can be meaningfully combined to construct a more comprehensive understanding of the developmental course of character virtues, such as hopeful future expectations, throughout childhood, adolescence, and early adulthood.

Hopeful future expectations are thought to be a key variable in character virtue development as one aspect of future orientation that helps to organize and energize behavior toward a productive adulthood (Schmid and Lopez 2011; Seligman et al. 2013). For example, hopeful future expectations was shown to moderate the role of intentional

self-regulation in promoting the "Five Cs" of positive youth development (Confidence, Competence, Character, Connection, and Caring; Schmid et al. 2011a, b).

We believe the conceptual importance attributed to this variable should be matched, at least, by an empirical description of the development of hopeful future expectations in the first three decades. Longitudinal data sets have previously provided only unconnected assessments of the developmental course of hopeful future expectations and its contribution to thriving. As such, there has been a discrepancy between theory and data in regard to this character virtue. Unfortunately, this discrepancy is not unique to this particular character virtue; much of the literature on character provides only a snapshot of developmental processes (Card in press).

We believe that it is unlikely that the resources will be available to conduct what would amount to multi-decade longitudinal studies—ones spanning the years prior to, during, and after the adolescence period—to generate the comprehensive information needed about either hopeful future expectations alone or, more generally, about the larger set of character virtues studied in the developmental science literature (e.g., Lerner and Callina 2014; Nucci in press). However, such new, expensive (in regard to both financial and human resources), and long-term research may not be needed if IDA techniques can be used to enhance the understanding of the developmental course of character virtues.

We believe that the findings presented here provide a proof of concept for the usefulness of this method to generate new and important information about hopeful future expectations. For instance, the present research provides the first evidence in the developmental science literature about the trajectory of hopeful future expectations, from ages 7 to 26. Thus, the use of IDA methods has revealed a heretofore unknown and theoretically unanticipated feature of the development of this character virtue: that is, the present study found that there was a decline in hopeful future expectations across ages 7 through 26. Without additional behavioral, emotional, or cognitive assessments of the participants in the data sets we used in our study, the reasons for the observed decline are uncertain. Nevertheless, we do have some hypotheses based on the future orientation literature that, given our surprising finding, would be exciting to test in future research about hopeful future expectations.

As an evaluation of the extent to which young people expect to realize future goals, hopeful future expectations may be guided by the same principles as optimism and hope, which are shown to flourish when individuals have high self-concept and can attribute achievements to stable, internal, and controllable successes (Nurmi 1991; Seligman et al. 2013). Thus, the decline in hopeful future expectations scores may reflect the fact that individuals are experiencing internal or external constraints as they approach the developmental tasks of adulthood, including educational attainment, vocational training, and establishing intimate partnerships (Nurmi 1991). Indeed, the study-based differences that we observed in our analyses may lend support to this hypothesis. The ACT participants had higher intercept scores than the referent group (CAMP study participants), whereas the 4-H and YES participants scored lower. One possible explanation for this finding is that many of the ACT participants were students at a post-secondary trade school that focuses on trade education and character development, and which boasts a nearly 100% job placement rate (Johnson et al. 2014).

Future applications of IDA could seek to examine additional predictors of hopeful future expectations trajectories. For instance, goal setting and management skills are thought to promote hopeful future expectations among young adolescents (Schmid et al. 2011b), and social connectedness (including trust and support from parents and peers) has been linked to the development of hope and hopeful future expectations from infancy through adolescence (Callina et al. 2014a; Erikson 1959; Snyder 1994). Moreover, some of these factors may be more salient for young people in out-of-school time activities that aim to promote positive character (Ettekal et al. 2015). Future research could examine additional sources of sampling heterogeneity that may be linked to hopeful future expectations, including participants' access to and participation in youth development programs.

Another predictor of decreasing hopeful future expectations scores may be simply life course and life-span development; perhaps hopeful future expectations scores decreased as participants approached adulthood, or perhaps they decreased due to structural, social, or cultural changes. Note that the Great Recession sparked by the United States housing crisis of 2007–2009 was the major historical event likely to affect the hopeful future expectations scores of our sample, and the average age of all participants during this event was about 16 years. However, our study was not able to account for differences due to cohort vs. age, and so we cannot attribute the decline in hopeful future expectations scores to maturation, on the one hand, or historical events, on the other (and it is likely that both are factors in hopeful future expectations). To investigate this issue more closely, we plotted the age of participants against the years of data collection, presented in Fig. 1. As we illustrated in the figure, we found that there are only two birth cohorts in our pooled sample. The first cohort is comprised of 4-H Study, ACT, and YES participants, who were born in the early- to mid-1990s. The second cohort is comprised of CAMP participants, who were born in the early- to mid-2000s. Future research could pool data from more diverse cohorts

to determine the extent to which the slope of hopeful future expectations scores is affected by cohort differences or maturation.

Moreover, we were also not able to examine a Growth Mixture Model, due to lack of covariance coverage (e.g., low percentage of participants who had data at two or more time points) in the hopeful future expectations items that were available across the time points and data sets. Growth Mixture Models would allow us to examine whether there are different trajectory profiles in our pooled sample. For example, we would hypothesize that there is greater variation in the hopeful future expectations scores of older participants, such that some participants have a hopeful future expectations profile that stays high and stable from childhood through young adulthood, whereas other participants have a profile that decreases following major life transitions, such as graduation from high school or college. We would also expect to find variation in the shape of individuals' trajectories, such that some youth experience a dip in hopeful future expectations during adolescence, whereas others show steady increasing, decreasing, or stable trajectories (indeed, these were the trajectories identified in a sample of 4-H Study youth ages 13–16; Callina et al. 2014a). For more information on the developmental course of future orientation in adolescence and early adulthood, we point readers to Nurmi, Steinberg, Trommsdorff, Seginer, and others (e.g., Nurmi 1991; Seginer 2009; Steinberg et al. 2009; Trommsdorff et al. 1979; Trommsdorff 1983).

The limitation of the covariance coverage of hopeful future expectations items available across studies and ages in our data set highlights a key issue for researchers who may be interested in designing studies that will eventually be useful for conducting IDA. Researchers should attempt to retain as many items as possible throughout the waves of data collection of any longitudinal study, particularly items like hopeful future expectations, which have good reliability and validity. Many of the items removed in our studies were done so to attenuate "survey fatigue" of the participants, and to make room for other constructs of interest.

However, a carefully considered measurement model and pilot testing of the survey instrument may help researchers avoid having to remove items later in the data collection. In addition, many cases were excluded from our analytic sample (almost half), in part because participants were missing data on some hopeful future expectations items and so did not have a hopeful future expectations scale score computed. The other reason participants were excluded from the analysis was due to missing age data. Researchers conducting longitudinal studies should make every attempt to collect complete information about age. Accurate age data can be difficult to obtain when collecting data from small children or from a large online sample; indeed, these were the challenges faced by the researchers who conducted the CAMP, YES, and 4-H studies, and the primary reasons why there was so much missing age data in the analyses presented here. However, researchers can introduce error, as well: when calculating age from date of birth and date of participation, for instance, researchers must take care to record when the data were collected for each participant.

Nevertheless, we see utility in our findings for practitioners, parents, and educators who seek to promote hopeful future expectations in youth, and for researchers who are involved in program evaluations of such efforts. In establishing both longitudinal and between-study invariance, we have demonstrated that hopeful future expectations may be a useful measure for diverse age groups and sampling designs. In addition, the decreasing hopeful future expectations trajectory identified here suggests that a successful youth development program should not be discouraged if it does not observe increasing hopeful future expectations scores in an evaluation. Stable hopeful future expectations scores observed over several time points may indicate that the program has a buffering effect on the general decline in hopeful future expectations scores among youth ages 7 through 26.

In addition to the rich ideas for building on the present substantive findings to further advance knowledge of the developmental course of hopeful future expectations and the importance of this development for positive youth development, the present research has implications for developmental science methodology, and for extension of our substantive and methodological work to the study of other character virtues. In essence, the research reported in this article is a proof-of-concept and illustrative example of the potential of IDA approaches for linking different data sets to describe the development of character virtues. With the growing interest in assessing character virtue development and character education programs, there is an increasing number of short-term longitudinal studies that provide depictions of changes in character virtues among specific groups of youth, studied with specific measures, across specific, age-limited portions of the life span (e.g., see Ettekal et al. 2015; Wang et al. 2016). Many of these studies have been supported by the same funder, the John Templeton Foundation, and thus are, at least in part, framed by Sir John Templeton's (1998, 2000) ideas about character virtues. In addition, and as illustrated by the four studies used in the present illustration of the potential of IDA, subsets of these investigations of character virtues development have been derived from the work of one research laboratory (here, the Institute for Applied Research in Youth Development at Tufts University).

Despite commonality of funding source or laboratory and an overall commitment to understanding character virtues development across substantial portions of the life span—in the case of the Institute for Applied Research in Youth

Development, the first three decades of life—the IDA approaches presented in this article represent the first empirical means to actually synthesize the information across different studies to assess whether an integrative account can be made about character virtues development. In the present study, we modeled a single hopeful future expectations trajectory from childhood through early adulthood and identified gender and race as significant predictors of the intercept for this trajectory. Although the individual studies used in this IDA demonstration contain relatively large sample sizes and span between 3 and 8 years of development, we were interested in modeling hopeful future expectations across a greater period of the life span. Future research could extend our demonstration to investigate further the psychological and contextual variables associated with the slope and intercept of developmental trajectories of hopeful future expectations. Of course, as we note again in this section, future IDA research should also involve the investigation of other character virtues.

There are also other potential uses of IDA we can envision. Theories framing contemporary developmental science—in particular, relational developmental systems theories (Lerner 2015; Overton 2015)—emphasize the non-ergodicity of human development (e.g., Mascolo and Fischer 2015; Molenaar and Nesselroade 2015; Raeff 2016; Rose 2015). Non-ergodicity refers to statistical methods that provide information about individuals' dynamic and diverse developmental trajectories (Molenaar and Nesselroade 2015). Therefore, by synthesizing different data sets pertinent to character virtues development, each providing information about the study of ontogenetic change across a portion of the life span, IDA approaches have the potential to depict the parameters of the specificity principle (Bornstein 2017) brought to the fore by a concern with the primacy of idiographic (non-ergodic) developmental trajectories. This principle points to the idiographic, as compared to the nomothetic, nature of human development. For instance, such integration may elucidate what specific character virtues, for individuals of what specific ages and other specific demographic attributes, developing in what specific settings, and having what specific experiences (e.g., participating, or not, in specific youth-development and/or character-education programs), and linked to what other specific individual and ecological variables, result in what specific developmental trajectories of both character virtues development and other theoretically-specified features of human development (e.g., active and positive civic engagement; Lerner et al. 2015) across what specific portions of the life course.

Such information would help enhance substantially the science of character virtues development and, as well, provide a cost-effective means to leverage the substantial investments of resources that have been made in the studies whose data are brought together through IDA approaches. Although the present research has focused on only one character virtue, hopeful future expectations, the success of this proof-of-concept assessment suggests that IDA represents a means to move the field to a more robust, empirical discussion of how extant research about character virtues development may be collaboratively synthesized.

For example, a network of character virtue development researchers might be created to systematically pursue such synthesis. Although participation in such a network would not preclude individual laboratories from conducting research pertinent to their own theory-guided questions about character virtues development, the benefit (or the virtue, if you will) of such a network would be that participants might plan their new research with an eye toward eventual integration through IDA approaches. New research could be launched with agreements to include some common items or measures of character virtues across studies. Similarly, studies could be planned so that there was some intentional age-period overlap across studies. One potential benefit of such a network would be in the diversity of participants of the pooled data. Such diversity could add to the knowledge base pertinent to the several specificity questions noted above.

Such advances in scholarship would serve not only the science of character virtues development but, as well, the needs of character educators, both in out-of-school-time programs and in school-based programs (Lerner, Vandell, & Tirrell, in press). At this writing, character educators have not had the information they knew they needed to adapt the knowledge base of character virtues development to fit the specific developmental niche of the individuals they serve (Berkowitz et al. 2016). A network of character virtues development scholars using IDA approaches would, then, be able to contribute substantially to this fundamental issue of application to educational settings.

Of course, the enthusiasm we clearly have about the potential value of future explorations and applications in research using IDA approaches must be tempered by the important limitations of the present research. Most fundamentally, the present proof-of-concept/illustration of the use of IDA approaches involves only one instance of a character virtue and only four studies, which were all conducted within the same lab (the Institute for Applied Research in Youth Development). Obviously, then, one next step is to expand the present demonstration to other character virtues. This extension can be done across additional studies beyond the four used in the present article and this work can occur within our own lab and, of course, within other labs that have multiple data sets. However, a second step—and one that will be a key to building the character virtues development network we have envisioned—would be to pool data from various researchers. In expanding the application

of IDA to incorporate new constructs and studies, we hope to also expand our understanding of the analytic techniques that make IDA possible; for instance, in the present study we used randomly generated data in our between-study invariance test. Further research about the method, advantages, and potential limitations of this technique is warranted.

Conclusion

The two purposes of the present research have been successfully met. Our research has provided heretofore unavailable information about the development of the character virtue of hopeful future expectations across a period of about 20 years, from middle childhood through early adulthood. We have discovered new and theoretically unexpected information about the developmental course of this character virtue: that is, hopeful future expectations appear to have a declining trajectory across these years. Our findings provide a rich basis of hypotheses that may now guide future research about this character virtue, which has been given the conceptual burden of being a key predictor of positive youth development. Developmental scientists—as well as practitioners, educators, parents, and mentors—should be aware that this virtue is a declining phenomenon during the adolescent period. Accordingly, the implications of our findings for diverse youth will be an important substantive focus for both the study of character development and the understanding of the bases of positive youth development.

In turn, the IDA approach presented in this article represents the first empirical attempt within the character virtue development literature at synthesizing information across different studies. Our substantive findings certainly justify continued exploration of the usefulness of this method. Our analyses provide evidence that IDA is a useful analytical tool for synthesizing data collected from multiple studies and a fruitful approach for character virtue development research.

Despite the limitations of the present research, we believe that the potential of IDA approaches to synthesize the extant knowledge base pertinent to character virtues development, and the additional potential of these approaches to provide a frame for the integration of future research, will fulfill the vision of researchers, practitioners, and funders (e.g., Templeton 1998). Our own hopeful expectation for the future is that the field of applied developmental science may be nearing a period of scientific collaboration wherein new and needed information about a core component of individual thriving and positive social relationships in civil society—character virtues—can be better described, explained, and optimized. Such scientific progress will enhance both individuals and diverse communities in which they live.

Acknowledgements The authors gratefully acknowledge the work of additional principal investigators, project directors, and other staff (including graduate and undergraduate students) to conduct the studies used in the present research. The authors also gratefully acknowledge the participation of schools and youth across the four studies, which made the present research possible. This research was supported in part by grants from the John Templeton Foundation, the Templeton Religion Trust, the National 4-H Council, and the Altria Corporation.

Author Contributions K.S.C. conceived of the study, spearheaded its design and coordination, prepared the data and performed and interpreted statistical analyses, and had primary responsibility for writing the full manuscript; S.K.J. conceived of the study, participated in the study design and coordination, performed and interpreted statistical analyses, and contributed to writing the manuscript; J.M.T. helped to prepare, manage, and analyze the data, and contributed to writing in the manuscript; M.B. contributed to writing the manuscript; M.B.W. contributed to writing the manuscript; and R.M.L. conceived of the study and contributed to writing the manuscript. All authors read and approved the final manuscript.

Conflict of Interest The authors declare that they have no competing interests.

Ethical Approval This research used secondary data based on studies conducted at the Institute for Applied Research in Youth Development, all of which received approval from the Internal Review Board at Tufts University.

Informed Consent All eligible students for the studies used in the current research were fully informed about their voluntary participation. If students wished to refrain from participation, or if their parents disagreed with their children's participation (as in the cases of the CAMP and 4-H studies), they were free to do so. Only students who provided permission or assent to participate, and who had parental permission to participate, were involved in the studies.

References

Bainter, S. A., & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development*, *16*, 1–10.

Berkowitz, M. W. (2011). What works in values education. *International Journal of Educational Research*, *50*(3), 153–158.

Berkowitz, M. W. (2012). Moral and character education. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook. vol. 2: Individual differences, cultural variations, and contextual factors in educational psychology* (pp. 247–264). Washington, DC: American Psychological Association.

Berkowitz, M. W., Bier, M. C., & McCauley, B. (2016). *Effective features and practices that support character development*. Paper presented at The National Academies of Sciences, Engineering, and Medicine Workshop on Approaches to the Development of Character, Washington, DC.

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

Bowers, E. P., Geldhof, G. J., Schmid, K. L., Napolitano, C. M., Minor, K., & Lerner, J. V. (2012). Relationships with important nonparental adults and positive youth development: An

examination of youth self-regulatory strengths as mediators. *Research in Human Development*, *9*(4), 298–316.

Bowers, E. P., Hilliard, L. J., Batanova, M., Stacey, D. C., Tirrell, J. M., Wartella, K., & Lerner, R. M. (2015). The arthur interactive media study: Initial findings from a cross-age peer mentoring and digital media-based character development program. *Journal of Youth Development*, *10*(3), 46–63.

Bowers, E. B., & Johnson, S. K. (2013). The adult social milieu and character development: Potential mechanisms of influence. In N. Hurd (Chair), *Mentoring and adolescent development: Assessing pathways and exploring complexities*. Symposium conducted at the annual meeting of the Association of Psychological Science, Washington, DC.

Callina, K. S., Johnson, S., Buckingham, M., & Lerner, R. M. (2014a). Hope in context: Developmental profiles of trust, hopeful future expectations, and civic engagement across adolescence. *Journal of Youth and Adolescence*, *43*(6), 869–883.

Callina, K. S., Mueller, M. K., Buckingham, M. H., & Gutierrez, A. S. (2014b). Building hope for positive youth development: Research, policy, and practice. In E. P. Bowers, G. J. Geldhof, S. K. Johnson, L. J. Hilliard, R. M. Hershberg, J. V. Lerner, & R. M. Lerner (Eds.), *Promoting Positive Youth Development: Lessons Learned from the 4-H Study* (pp. 71–94). New York, NY: Springer.

Callina, K. S., Mueller, M. K., Napolitano, C. M., Lerner, J. V., & Lerner, R. M. (2016). Positive youth development: Relational developmental systems approaches to thriving from childhood to early adulthood. In S. J. Lopez, L. M. Edwards, & S. C. Marques (Eds.), *The Oxford handbook of positive psychology*. 3rd edn New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199396511.013.11.

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.

Callina, K. S., Snow, N., & Murray, E. D. (in press). The history of philosophical and psychological perspectives on hope: Toward defining hope for the science of positive human development. In M. Gallagher & S. J. Lopez (Eds.), *The Oxford Handbook of Hope*. New York, NY: Oxford University Press.

Card, N. A. (2011). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

Card, N. A. (in press). Methodological issues in measuring the development of character. *Journal of Moral Education*.

Carver, C. S., & Scheier, M. F. (2002). The hopeful optimist. *Psychological Inquiry*, *13*(4), 288–290.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data ver- sus aggregated data. *Psychological Methods*, *14*, 165–176.

Cumming, G. (2013). *The new statistics: Estimation for better research*. Retrieved from www.thenewstatistics.com.

Curran, P. J. (2009). The seemingly quixotic pursuit of a cumulative psychological science: Introduction to the Special Issue. *Psychological Methods*, *14*(2), 77–80.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychology Methods*, *14*, 81–100.

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*, 365–380.

DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, *46*(1), 137–149.

Erikson, E. H. (1959). Identity and the life cycle. *Psychological Issues*, *1*, 50–100.

Ettekal, A. V., Callina, K. S., & Lerner, R. M. (2015). The promotion of character through youth development programs. *Journal of Youth Development*, *10*(3), 6–13.

Ferris, K. A., Vest Ettekal, A., Agans, J. P., Burkhard, B. M., & Lerner, R. M. (2016). Character development through youth sport: High school coaches' perspectives about a character-based education program. *Journal of Youth Development*, *10*(3), 127–140.

Geldhof, G. J., Bowers, E. P., & Lerner, R. M. (2013). Special section introduction: Thriving in context: Findings from the 4-H study of positive youth development. *Journal of Youth and Adolescence*, *42*(1), 1–5.

Geldhof, G. J., Pornprasertmanit, S., Schoemann, A., & Little, T. D. (2013). Orthogonalizing through residual centering: Applications and caveats. *Educational and Psychological Measurement*, *73*, 27–46.

Geldhof, G. J., Porter, T., Weiner, M. B., Malin, H., Bronk, K. C., Agans, J. P., Mueller, M., Damon, W., & Lerner, R. M. (2014). Fostering youth entrepreneurship: Preliminary findings from the Young Entrepreneurs Study. *Journal of Research on Adolescence*, *24*, 431–446.

Grimm, K. J., & Ram, N. (2012). Growth curve modeling from a structural equation modeling perspective. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 411–431). New York, NY: Guilford Press.

Hilliard, L. J., Hershberg, R. M., Wang, J., Bowers, E. P., Chase, P. A., Champine, R. B., & Lerner, R. M. (2014). Program innovations and character in Cub Scouts: Findings from year 1 of a mixed-methods, longitudinal study. *Journal of Youth Development*, *9*(4), 4–30.

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, *14*, 150–164.

Johnson, S. K., Hershberg, R. M., Arbeit, M. R., DeSouza, L. M., Callina, K. S., & Gutierrez, A. S., et al. (2014). Exploring characteristics of young adult men: Initial findings from a mixed methods evaluation of an all-male character-focused trade school. *Journal of Character Education*, *10*(2), 129–154.

Lapsley, D. K., & Narvaez, D. (2006). Character education. In A. Renninger, & I. Siegel (Eds.), *Child psychology in practice*. Volume 4 of the Handbook of Child Psychology. 6th edn. (pp. 248–296). New York: Wiley.

Lerner, R. M. (2015). Preface. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science*. 7th edn. (pp. xv–xxi). Hoboken, NJ: Wiley.

Lerner, R. M., & Callina, K. S. (2014). Relational developmental systems theories and the ecological validity of experimental designs. *Human Development*, *56*(6), 372–380.

Lerner, R. M., Lerner, J. V., Almerigi, J., Theokas, C., Phelps, E., & Gestsdóttir, S., et al. (2005). Positive youth development, participation in community youth development programs, and community contributions of fifth-grade adolescents: Findings from the first wave of the 4-H study of positive youth development. *Journal of Early Adolescence*, *25*(1), 17–71.

Lerner, R. M., Lerner, J. V., Bowers, E., & Geldhof, G. J. (2015). Positive youth development and relational developmental systems. In W. F. Overton, & P. C. Molenaar (Eds.), *Theory and Method. Volume 1 of the Handbook of child psychology and developmental science*. 7th edn. (pp. 607–651). Hoboken, NJ: Wiley.

Lerner, R. M., Schwartz, S. J., & Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human development:

Theoretical and methodological issues. *Human Development*, *52* (1), 44–68.

Lerner, R. M., Vandell, D. L., & Tirrell, J. M. (in press). Approaches to the development of character: Meanings, methods, and measures. *Journal of Character Education*.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.

Lopez, S. J., Rose, S., Robinson, C., Marques, S. C., & Pais-Ribeiro, J. L. (2009). Measuring and promoting hope in school children. In R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.), *Handbook of positive psychology in the schools* (pp. 37–51). Mahwah, NJ: Lawrence Erlbaum.

Markus, H., & Nurius, P. (1986). Possible selves. *The American Psychologist*, *41*(9), 954–969.

Mascolo, M. F., & Fischer, K. W. (2015). Dynamic development of thinking, feeling, and acting. In W. F. Overton, & P. C. Molenaar (Eds.), *Theory and method*. Volume 1 of the *Handbook of Child Psychology and Developmental Science*. 7th edn. (pp. 113–161). Hoboken, NJ: Wiley.

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

McArdle, J. J., & Nesselroade, J. R. (2003). Growth cure analyses in contemporary psychological research. In S. H. Cohen, W. Velicer (Eds.), *Comprehensive handbook of psychology: Vol. 2 Research methods in psychology* (pp. 447–480). New York: Pergamon Press.

Molenaar, P. C. M., & Nesselroade, J. R. (2015). Systems methods for developmental research. In W. F. Overton, & P. C. Molenaar (Eds.), *Handbook of child psychology and developmental science. Vol. 1: Theory and method*. 7th edn. (pp. 652–682). Hoboken, NJ: Wiley.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide*. 6th edn. Los Angeles, CA: Muthén & Muthén.

Nurmi, J. E. (1991). How do adolescents see their future? A review of the development of future orientation and planning. *Developmental Review*, *11*, 1–59.

Nucci (in press). Character: A multi-faceted developmental system. *Journal of Character Education*.

Overton, W. F. (2015). Process and relational developmental systems. In W. F. Overton, & P. C. Molenaar (Eds.), *Handbook of child psychology and developmental science. Vol. 1: Theory and method*. 7th edn. (pp. 9–62). Hoboken, NJ: Wiley.

Raeff, C. (2016). *Exploring the dynamics of human development: An integrative approach*. London, England: Oxford University Press.

Rose, T. (2015). *The end of average: How we succeed in a world that values sameness*. New York, NY: HarperCollins Publishers.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, *64*(2), 92–107.

Schmid, K. L., & Lopez, S. (2011). Positive pathways to adulthood: The role of hope in adolescents' constructions of their futures. In R. M. Lerner, J. V. Lerner, & J. B. Benson (Eds.), *Advances in child development and behavior: Positive youth development* 41, (72–89). London, England: Academic Press.

Schmid, K. L., Phelps, E., & Lerner, R. M. (2011a). Constructing positive futures: Modeling the relationship between adolescents' hopeful future expectations and intentional self-regulation in predicting positive youth development. *Journal of Adolescence*, *34*(6), 1127–1135.

Schmid, K. L., Phelps, E., Kiely, M. K., Napolitano, C. M., Boyd, M. J., & Lerner, R. M. (2011b). The role of adolescents' hopeful futures in predicting positive and negative developmental trajectories: Findings from the 4-H study of positive youth development. *Journal of Positive Psychology*, *6*(1), 45–56.

Seginer, R. (2009). *Future orientation: Developmental and ecological perspectives*. New York: Springer Series.

Seligman, M. E. P. (1991). *Learned optimism: How to change your mind and your life*. New York: Random House.

Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8*(2), 119–141.

Shrout, P. E. (2009). Short and long views of integrative data analysis: comments on contributions to the special issue. *Psychological Methods*, *14*(2), 177–181.

Snyder, C. R. (1994). *The psychology of hope: You can get there from here*. New York: Free Press.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., Gibb, J., Langelle, C., & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology*, *60*(4), 570–585.

Sokol, B. W., Hammond, S. I., & Berkowitz, M. W. (2010). The developmental contours of character. In T. Lovat, R. Toomey, N. Clement (Eds.), *International research handbook on values education and student wellbeing* (pp. 579–603). Dordrecht, Netherlands: Springer.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Development*, *80*(1), 28–44.

Templeton, J. M. (1998). *The humble approach: Scientists discover God*. Radnor, PA: Templeton Press.

Templeton, J. M. (2000). *Worldwide laws of life: 200 Eternal spiritual principles*. Radnor, PA: Templeton Press.

Trommsdorff, G. (1983). Future orientation and socialization. *International Journal of Psychology*, *18*, 381–406.

Trommsdorff, G., Lamm, H., & Schmidt, R. W. (1979). A longitudinal study of adolescents' future orientation (time perspective). *Journal of Youth and Adolescence*, *8*(2), 131–147.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492.

Wang, J., Batanova, M., Ferris, K. A., & Lerner, R. M. (2016). Character development within the relational developmental systems metamodel: A view of the issues. *Research in Human Development*, *13*(2), 91–96.

Wang, J., Ferris, K. A., Hershberg, R. M., & Lerner, R. M. (2015). Developmental trajectories of youth character: A five-wave longitudinal study of cub scouts and non-scout boys. *Journal of Youth and Adolescence*, *44*, 2359–2373.

Widaman, K. F., Grimm, K. J., Early, D. R., Robins, R. W., & Conger, R. D. (2013). Investigating factorial invariance of latent variables across populations when manifest variables are missing completely. *Structural Equation Modeling*, *20*, 384–408.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*, 16–37.

Kristina Schmid Callina is a Research Assistant Professor at the Institute for Applied Research in Youth Development at Tufts University. Her research interest is on character virtue development from childhood through early adulthood. In particular, she studies the role of hope in positive youth development, including the psychosocial strengths that contribute to hope (e.g., optimism and self-regulation) as well as assets in the context, such as positive youth-adult relationships, that promote it.

Sara K. Johnson is an Assistant Professor in the Eliot-Pearson Department of Child Study and Human Development at Tufts University. Her substantive research interests within the field of

positive youth development focus on youth contribution and civic engagement and their relationship with identity development. She is also interested in quantitative research methods to identify population heterogeneity, such as a latent class and latent profile analyses.

Jonathan M. Tirrell is a PhD candidate at the Institute for Applied Research in Youth Development at Tufts University. His research interests include moral and character development and, in particular, the role of forgiveness and related concepts in promoting thriving and social justice.

Milena Batanova is a Research Assistant Professor at the Institute for Applied Research in Youth Development at Tufts University. Her research interests include empathy development, aggression and bullying prevention, and bystander intervention. In particular, she is interested in the varied meanings, antecedents, and outcomes of empathy for promoting positive youth development, and the roles of ecological assets, like schools and families, in helping to cultivate empathy and related attributes in youth (and adults).

Michelle B. Weiner is a Research Assistant Professor at in the Lynch School of Education at Boston College. Her research interests include the development of skills related to entrepreneurship and character virtue development. In particular, she is interested in the importance of goal-setting skills among youth for promoting positive youth development.

Richard M. Lerner is the Bergstrom Chair in Applied Developmental Science, Professor of Child Study and Human Development, and Director of the Institute for Applied Research in Youth Development at Tufts University. He received his Ph.D. in developmental psychology from the City University of New York. His work integrates the study of public policies and community-based programs with the promotion of positive youth development and youth contributions to civil society.