INT'L. J. AGING AND HUMAN DEVELOPMENT, Vol. 68(4) 289-320, 2009

THE WISDOM DEVELOPMENT SCALE: FURTHER VALIDITY INVESTIGATIONS

JEFFREY A. GREENE, PH.D.
University of North Carolina at Chapel Hill

SCOTT C. BROWN, PH.D.
Colgate University

ABSTRACT

Researchers are gaining an interest in the concept of wisdom, a more holistic yet often ineffable educational outcome. Models of wisdom abound, but few have rigorously tested measures. This study looks at Brown's (2004a, 2004b) Model of Wisdom Development and its associated measure, the Wisdom Development Scale (WDS; Brown & Greene, 2006). The construct validity, measurement invariance, criterion validity, and reliability of scores from the WDS were assessed with over 3000 participants from two separate groups: one a sample of professionals and the other a sample of college students. Support for construct validity and reliability with these samples was found, along with measurement invariance. Latent means analyses showed predicted discrimination between the groups, and criterion validity evidence, with another measure of collegiate educational outcomes, was found.

As societies and societal institutions continue to struggle to define themselves in a new era of standards and empirically based decision-making (National Research Council, 2002), there are those who are seeking to supplement the common metrics with more integrative measures of personal growth. Indeed, the positive psychology movement (Seligman & Csikszentmihalyi, 2000) stresses that deficit-based understandings of human functioning ignore the many ways

doi: 10.2190/AG.68.4.b
http://baywood.com

people circumvent the difficulties in their lives to find success. One positive psychology construct that is garnering more interest in the fields of development, education, and psychology is wisdom (Ardelt, 2003; Baltes & Smith, 1990, 2008; Sternberg, 1985; Webster, 2003, 2007; Wink & Helson, 1997). There are numerous models of wisdom, but most represent an attempt to characterize and understand the cognitive, and more recently the affective, social, and moral qualities that characterize those who are commonly considered wise (Ardelt, 2004; Sternberg, 2003). It is hoped that our societal institutions promote wisdom, as well as performance. However, the case for making wisdom a priority outcome in this era of standards has been made all the more difficult due to problems with its measurement.

Capturing wisdom and its development is an enormously challenging task (Kunzmann & Baltes, 2003). Scores from psychometric instruments must be carefully scrutinized to ensure that they are adequate indicators of wisdom. Such evidence includes the reliability of the scores as well as their content validity (i.e., their ability to capture the breadth and depth of wisdom as conceptualized in theoretical models) and construct validity (i.e., the degree to which the items posited to measure the construct sufficiently capture that construct while not being influenced by any irrelevant sources of variance; see Messick, 1989, for a review of these issues). Numerous measures of wisdom have been advanced, but unfortunately they have often had low levels of reliability (Ardelt, 1997) and questionable content validity (Baltes & Smith, 1990), or have not been tested with large enough sample sizes to adequately test the construct validity of their scores (Webster, 2003, 2007).

Brown and Greene (2006) have produced a measure of wisdom based on Brown's (2004b) Model of Wisdom Development. An initial study (Brown & Greene, 2006) demonstrated the reliability and construct validity of scores from the Wisdom Development Scale (WDS) with a collegiate sample, but cross-validation of those findings with other samples, as well as an examination of other types of validity, such as predictive and criterion-based studies, are needed. A rigorously tested, effective measure of wisdom could be used to understand the development of the construct over time as well as how it can be influenced through various types of interventions and experiences. This understanding could then be used to help justify allocating resources toward the facilitation of wisdom development and allow stakeholders to assess the influence of those interventions and experiences upon cognitive, affective, social, and moral growth.

EMPIRICAL RESEARCH REGARDING WISDOM

One of the first questions addressed in empirical research on wisdom is whether individuals see the construct as separate and distinct from other qualities such as intelligence. Both survey and interview data indicate that laypeople do view wisdom as a distinct construct that overlaps with, but is distinct from, intelligence

(Holliday & Chandler, 1986; Sternberg, 1985). Given the uniqueness of the construct, researchers have since been investigating means of measuring wisdom and its development. Two of the more predominant means of assessing wisdom include wisdom-related performance and latent variable measures.

Researchers at the Max Planck Institute in Berlin, Germany have conceptualized wisdom as a compendium of expert knowledge regarding the "fundamental pragmatics of life" (Baltes & Smith, 2008, p. 58) that affords good judgment regarding life matters that are both uncertain and important (Baltes & Smith, 1990; Baltes & Staudinger, 2000). Rather than using self-report measures to capture this kind of knowledge, these researchers have presented participants with wisdom-related performance measures, such as asking them to respond to problems concerning morality and life-planning. Thus, participants' performance in these situations is determined by rating the wisdom of their responses, and used as an indicator of the amount of expert knowledge the individual holds. These ratings have been performed by both trained experts as well as individuals who hold only implicit, or folk, theories of wisdom. These ratings have been reliable across these groups (Baltes & Smith, 2008). Their findings indicated a weak developmental trend that they interpreted as support for their model: younger participants displayed wiser performance in areas in which they had more expertise, whereas older participants were more likely to display wise performance in familiar and unfamiliar scenarios. However, age alone was not a reliable predictor of wisdom. Rather, wisdom-related knowledge and judgment skills came about as a function of psychological, social, work-related, and historical experiences (Baltes & Smith, 2008). Intelligence was a stronger predictor of wisdom in adolescents than it was in adults. Likewise, individuals hypothesized to be more likely to have wisdom, such as clinical psychologists, consistently evidenced wiser performance than laypeople (Baltes & Smith, 2008; Smith & Baltes, 1990; Staudinger, Maciel, Smith, & Baltes, 1998).

While the wisdom-related performance literature is both vast and compelling, there are concerns that the measures better capture participants' ability to postulate about wise performance, rather than assess their wisdom per se (Ardelt, 2004; Webster, 2003). Researchers interested in assessing wisdom as a latent construct have utilized quantitative survey methods, which require instruments whose scores show strong reliability and validity (DeVellis, 2003; Kline, 2005). Ardelt (2003) created and tested a wisdom scale covering three dimensions: cognition, affect, and reflection. Utilizing confirmatory factor analyses, scores from her scale showed good construct, predictive, and discriminant validity. However, a review of the literature found no further research into this scale, or whether its psychometric qualities were confirmed across multiple samples and subpopulations.

Webster (2003, 2007) created and tested a wisdom scale comprised of five dimensions: critical life experience, humor, openness, reminiscence, and emotional regulation. Initial exploratory factor analyses provided some support for the

scale's psychometric properties, and a follow-up study published in 2004 utilized confirmatory methods with a new sample. While determining the appropriate statistical criteria for accepting data as a good fit to the underlying theoretical model is a contentious area of debate (cf. Hayduk & Glaser, 2000), Webster's findings do not meet the criteria recommended by Hu and Bentler (1999), suggesting that the instrument requires revision before being used as a measure of wisdom.

In general, Ardelt (2004) has argued that the definition of wisdom remains elusive, with a common understanding that it must be multidimensional. Empirical research into the construct has been promising, but concerns remain regarding wisdom-related performance methods, as well as the psychometric qualities of survey instruments (Brown & Greene, 2006). Brown (2004b) has asserted a new multidimensional model of wisdom development, and chosen to measure the constructs using a survey-based instrument. As this study was designed to explore this theoretical framework and instrument, Brown's model is described next.

THE WISDOM DEVELOPMENT SCALE: THEORY AND TESTING

Theory

This study is based on Brown's (2004b) Model of Wisdom Development, a framework that describes wisdom, how wisdom develops, and the conditions that facilitate the development of wisdom (see Figure 1). While originally conceptualized within the realm of education, the model can be extrapolated to individuals both inside and outside of traditional educational systems. In Brown's model, wisdom is comprised of six interrelated factors or dimensions: Self-Knowledge, Understanding of Others, Judgment, Life Knowledge, Life Skills, and Willingness to Learn.

Self-Knowledge describes how well a person knows his or her own interests, strengths, weaknesses, and values. Self-Knowledge is characterized by personal authenticity and genuineness kept constant in a variety of contexts, and an internal locus of success/fulfillment/satisfaction in regards to their relationships and goals. *Understanding of Others* describes a person's deep understanding of a wide variety of people in varying contexts, a genuine interest in learning about others (attentiveness, empathy), the capability of engaging them (various approaches), a willingness to help them, and possession of advanced communication skills that enable one to articulate thoughts in a way meaningful to another person.

Judgment refers to the knowledge that there are different ways of looking at an issue when making key decisions, and that one must take into account a variety of viewpoints, the past, and the present context, as well as one's own background influences. Judgment is characterized by acuteness of perception and discernment. *Life Knowledge* includes recognition of the interconnectedness

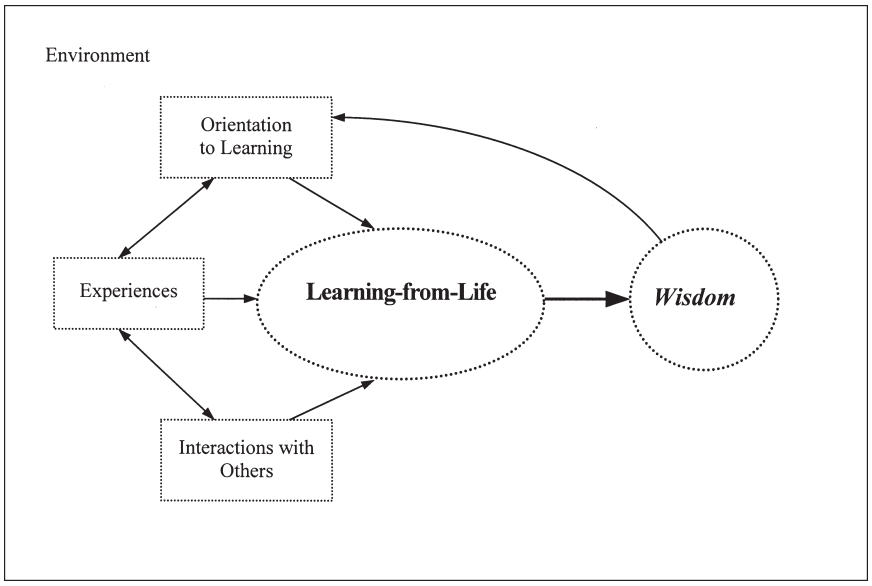


Figure 1. Brown's model of wisdom development.

between people and the natural world, knowledge and ideas, and the ability to look at the deeper meanings and questions in life. Life Knowledge is characterized by a capacity to grasp the central issue, find one's way in a time of darkness, and understand the realities and uncertainties of life, over the life span.

Life Skills includes the ability to manage one's daily multiple roles and responsibilities effectively. Life Skills is practical competence, an ability to understand systems and anticipate problems, with tools and strategies for dealing with multiple contexts in life. *Willingness to Learn* describes a basic humility in what one knows and continual interest in learning about the world.

Development. Wisdom develops when people go through the key "learning-from-life" process, where they reflect, integrate, and apply the lessons that they have learned, in and out of class, on and off campus, to their lives. The three conditions that directly facilitate the development of wisdom are a person's *orientation to learning, experiences*, and *interactions with others*. These conditions all take place in a particular environment, with a context that influences a person's orientation to learning and development. This context furnishes experiences, as do the people within that setting. *Orientation to Learning* refers to the attitude, level of engagement and potential for gaining knowledge when one interfaces with activities and people. This can include a general orientation to life, or vary by specific areas or situations, in addition to the person's past

as it comes to bear on any new interactions. *Experiences* include any activity, structured and unstructured. *Interactions with Others* includes all general experiences with others, experiences with people different from oneself, and in particular relationships such as friendships, family, and experiences with influential people. *Environment* refers to general settings, and provides the context where one's orientation to learning, variety of experiences, and interactions with people interact in various combinations to produce wisdom. Brown's Model of Wisdom Development has been the basis for research on post-college decision making (Brown, 2004a), and has been used as a framework to guide policies and practice in educational settings (Brown, 2002a, 2002b, 2006).

Comparing Brown's Theory to Other Models of Wisdom

Brown's theory has many connections to the professional literature on wisdom. Although Brown's six original dimensions of wisdom bear varying degrees of similarity to previous conceptions of wisdom, the definition of wisdom itself is generally more expansive. Other scholars who have researched wisdom, when discussed as a group, have created constructs that correspond with every one of Brown's dimensions of wisdom except *willingness to learn.* For example, Holliday and Chandler (1986) did a principal components analysis that yielded five factors related to wisdom: exceptional understanding of essences and contexts (akin to Brown's *life knowledge*) and the self (*self knowledge*), judgment (*judgment*), general competencies (*life skills*), and interpersonal skills and social unobtrusiveness (*understanding of others*). Sternberg's (1990) research on implicit theories of wisdom and their relationship to intelligence and creativity generated a conception of wisdom similar to Brown's theory: a deep understanding of self and others (*self-knowledge, understanding of others*), expeditious use of information, ability to learn from ideas and environment, perspicacity, discernability, and judgment (*judgment*). Schuman's (1982) conception of wisdom is to refine an individual's sense of a "rightfittingness" in the world (*self-knowledge*). Although these scholars' work might approximate wisdom as it is conceptualized in Brown's theory, they do not necessarily cover all aspects of each dimension, with similarities in breadth, but not necessarily in depth. Additionally, Brown's theory not only defines wisdom, but also illustrates how it develops and what influences its development.

Previous Testing of the WDS

In an initial study regarding the reliability and validity of scores from the WDS, Brown and Greene (2006) gathered data from 1188 college students. Utilizing an exploratory factor analysis to identify subscales and then a confirmatory factor analysis on a validation sample to confirm those subscales, Brown and Greene (2006) demonstrated evidence of construct validity of scores for five of

the six subscales measuring the factors in Brown's (2004b) Wisdom model. They did not find support for the Willingness to Learn factor subscale. Brown and Greene (2006) also found that two of the posited factors for the wisdom model were measured using two separate subscales each (see Table 1). This study provided valuable initial results regarding the construct validity of scores from the WDS, but further testing with different populations is needed before the instrument can be recommended for use in applied settings.

Purpose of This Study

Wisdom is a difficult thing to define, let alone measure. While Brown and Greene's (2006) study was an important first step, this study presents new and more varied evidence for the validity of scores from the WDS, and by extension Brown's (2004b) model. In general, the argument for an instrument's validity is stronger when the statistical analysis is confirmatory, rather than exploratory (DeVellis, 2003), as confirmatory analyses are more rigorous. It was our goal to gather a large enough sample to allow a strong test of the validity of scores from the WDS. We gathered over 3000 participants. Likewise, validity is dependent upon scores derived from a certain sample at a certain time, so any argument for validity must include multiple administrations drawn from different populations. Finally, an argument for validity is strongest when multiple kinds of validity evidence are presented (Messick, 1989). In this study, four kinds of validity evidence were sought: construct validity, measurement invariance, discriminant validity, and criterion validity.

Table 1. Dimensions of the Wisdom Development Scale (WDS)

Original	Wisdom Development Scale
Self-Knowledge	Self-Knowledge
Interpersonal Understanding	Altruism Leadership
Judgment	Judgment
Life Knowledge	Life Knowledge
Life Skills	Life Skills Emotional Management
Willingness to Learn	Willingness to Learn

Construct Validity and Measurement Invariance

With this study we hoped to expand upon the results of Brown and Greene (2006) by examining the construct validity of scores from the WDS with two new samples, one including adult higher education professionals and another comprised of college students. Our hypotheses were that scores from the WDS, individually in both samples, would have acceptable data-model fit using Hu and Bentler's (1999) criteria for confirmatory factor analyses (CFA). Despite the failure to find evidence of a coherent Willingness to Learn factor in Brown and Greene (2006), we also hypothesized that the Willingness to Learn factor would be supported in these samples. Next, we hypothesized that the WDS would display acceptable data-model fit using a multiple-groups analysis that evaluated both samples simultaneously. Finally, we hypothesized strong measurement invariance (Meredith, 1993) of the factor loadings across samples, providing further evidence of construct validity. The variance extracted from each item was also used to assess construct validity, with a value of .5 or higher being the goal (Gorsuch, 1983).

Discriminant Validity

In addition, we hypothesized that latent mean scores from the WDS subscales would show evidence of discriminant validity between the two samples. According to Brown (2004b), while wisdom does not have a perfect relationship with age, it should be the case that older populations, such as the professional sample, have higher latent mean scores on the WDS subscales than younger populations, such as the collegiate sample. Evidence supporting this hypothesis would provide further support for the construct validity of scores from the WDS.

Criterion Validity

Finally, we wished to test the criterion or concurrent validity of scores from the WDS. We could not find other instruments of good quality that measured wisdom in the way it is conceptualized by Brown (2004b). However, Brown's Model of Wisdom Development has similarities to the work of Chickering and Reisser (1993) on college student development. Chickering's theory is based on seven vectors that begin with the first three stages happening simultaneously:

1. Achieving Competence (intellectual, physical, and interpersonal);
2. Managing Emotions (awareness and integration of emotions); and
3. Moving through Autonomy toward Interdependence.

This leads to the critical fourth stage:

4. Developing Mature Interpersonal Relationships (tolerance of interpersonal and intercultural differences, capacity for intimacy).

The remaining stages can develop simultaneously:

5. Establishing Identity (comfort with appearance, sexual orientation, and "who I am" issues);
6. Developing Purpose (assess and integrate vocational plans, personal interests, and interpersonal commitments); and
7. Developing Identity (clarification of personal beliefs that are congruent with one's own values).

We found that the Iowa Student Development Inventories (Hood, 1997), which measure aspects of Chickering's theory of student development, contained three subscales we expected to correlate with subscales of the WDS.

The reliability and validity of scores from the Iowa scales have been tested in numerous settings with numerous college student populations (see Hood, 1997). Given that the Iowa scales were developed for use with college students, we utilized them with the college student sample only. We had hoped to find other scales that would naturally correlate with the remaining subscales in the WDS, but chose to pursue this in the future given that the number of items between the WDS and the Iowa scales numbered over 100, and we were wary of participant fatigue. We hypothesized that the Iowa scales of developing autonomy, developing purpose, and managing emotions would correlate with the WDS subscales of altruism, life skills, and emotional management, respectively.

Reliability

Finally, we were also interested in the reliability of scores from the WDS. Thus, reliabilities of the latent factor scores used in the construct validity analysis were computed. In addition, the reliabilities of the summed scores in the criterion validity analyses were also computed. We hypothesized that each of these measures would be above .7, a common metric in the field (Cronbach & Shavelson, 2004).

METHOD

This study was approved by the sponsoring institution's Institutional Review Board. The instruments were administered electronically through e-mail solicitation. We attempted to follow the suggestions of Cook and colleagues (2000) who found that multiple, but not too many, contacts and incentives often lead to higher response rates. E-mail administration is both more efficient and less personally persuasive than face-to-face solicitation, but we hoped that our invitation to participate was compelling. Finally, we sampled large numbers of people in the hope of getting enough participants to meet the CFA sample size recommendations of Kline (2005) with the knowledge that electronic solicitation of participation often leads to lower response rates (Cook et al., 2000).

Participants

Professional Sample

We received permission to send invitation e-mails to the entire population of a major professional organization serving the needs of college administrators and professionals. This e-mail invitation stated that we were interested in surveying the professionals' "perspectives and attitudes on life." The survey was described as having 84 items and taking 10 to 15 minutes to complete. Twelve Amazon.com gift certificates were used as incentives with a random drawing from the pool of completed surveys used to determine the winners. A hyperlink to the survey was provided and participants had to login using their email address.

Of the 6830 e-mails successfully delivered, there were 2715 completed surveys for a response rate of 40%. This response rate seems to be in line with other electronic survey administrations (Cook et al., 2000). The respondents' mean age was 34.1 years with a standard deviation of 10.5. The majority of respondents were white (83.1%), followed by African-American (6.9%), Multi-racial (3.1%), Latino/a (3.1%), Asian-American (2.6%), International (.7%), and Native-American professionals (.4%). One thousand eight hundred and twelve (66.8%) of the respondents were female and 893 (32.9%) were male. No respondents identified as transgendered.

College Student Sample

For the college student administration, 3000 e-mail addresses were randomly selected from the entire undergraduate and graduate population of a major mid-Atlantic University. These students received an e-mail invitation to participate in a survey "to help us understand students' perspectives and attitudes on life." The e-mail listed the length of the survey (171 items, approximately 25-30 minutes to complete) as well as the incentives, which in this case were Amazon.com gift certificates to be awarded randomly to five students who completed the survey. Respondents logged in following the same process as the professional sample.

Of the 3000 e-mails, 25 were not delivered successfully. Three hundred and thirty-eight valid responses were received after checking for duplicate e-mails and submissions, for a response rate of 11.4%. We had hoped for a higher response rate but surmised that the length of the survey (171 items) dissuaded many students from participating, despite the findings of Cook and colleagues (2000). The respondents' mean age was 21.2 with a standard deviation of 10.5. There were 93 freshmen (27.5%), 71 sophomores (21%), 80 juniors (23.7%), 76 seniors (22.5%), and 17 graduate students (5%). Two hundred and one (59.5%) of the participants were women, 137 (40.5%) were men, and none identified as transgendered. The majority of respondents were white (71.9%), followed by Asian-American students (12.7%), Multi-racial students (4.4%), African-American students (3.8%), Latino/a students (3.3%), International students (2.7%), and

Native-American students (.9%). The University purports to have an even split between men and women, with a 32% minority undergraduate population and a 17.5% minority graduate population, so the sample is somewhat skewed in terms of sex but similar to the University population in terms of majority versus minority race populations. Thus, in these ways this sample appeared to be representative of the population, a quality that may be more important than response rate (Cook et al., 2000).

Measures

Wisdom Development Scale

The WDS is an instrument measuring eight factors with 79 items. The original theory contained five factors: Self-Knowledge, Interpersonal Understanding, Judgment, Life Knowledge, Life Skills, and Willingness to Learn. Analyses of the WDS revealed that two factors were best captured using subscales (Brown & Greene, 2006). Interpersonal Understanding divided into two subscales: Altruism and Inspirational Engagement. Life Skills also divided into two subscales: Life Skills and Emotional Management. Responses were gathered using a Likert-style scale ranging from one to seven with one being strongly disagree and seven being strongly agree. All of the items in the version of the WDS used in Brown and Greene (2006) were worded positively. For this study eight of the items were worded negatively and thus hypothesized to load negatively on their respective factors. See Appendix A for sample items. Previous Cronbach's α for the individual subscales ranged from .84 to .88 and subscale intercorrelations ranged from .43 to .86.

Iowa Student Development Inventories

The Iowa Student Development Inventories (Hood, 1997) include numerous subscales measuring six aspects of Chickering and Reisser's (1993) theory of college student development. Three of these subscales were hypothesized to correlate with three subscales from the WDS. The Iowa Developing Autonomy scale contained a subscale of interdependence that measures "interdependence with others—neither totally independent nor totally dependent" (Hood, 1997, p. 34) and has 15 items. These items include "I feel I have a lot to contribute to my school or community" and "I think we should share our wealth and expertise with poor countries" (Hood, 1997, p. 47). Scores from this subscale have a reported reliability of .80. We hypothesized that this subscale would correlate with the WDS scale Altruism, which includes items such as "I am sensitive to the needs of others" and "I use my influence for the good of others."

The Iowa Developing Purpose subscale entitled "Style of Life" measures how much "clarity in the type of life [college students] wish to lead" (Hood, 1997, p. 79) and has 15 items. Items from this subscale include "I feel confident I

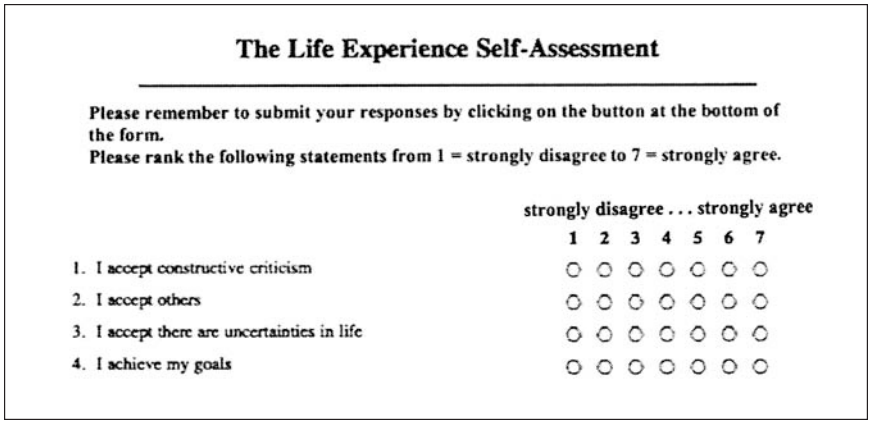
know where I am going in my life" and "I think about how my personal values relate to my career plans" (Hood, 1997, p. 91). Scores from this subscale have reported reliabilities ranging from .69 to .87. This scale seemed quite similar to the WDS subscale of Life Skills, which contains items such as "I have a sense of purpose in my life" and "I achieve my goals," thus we hypothesized a strong correlation between the two.

The Iowa Managing Emotions scale has 60 items and measures "awareness of and integration of emotions" (Hood, 1997, p. 22), with scores having a reported reliability of .90. Items from this scale include "When feeling frustrated, I find a solution and move on to other tasks" and "I am conscious of what makes me happy." We hypothesized scores from this subscale would correlate with the WDS Emotional Management subscale, which includes items such as "I manage stress effectively" and "I manage my emotions effectively."

Procedure

As stated above, participants received an e-mail inviting them to participate by clicking on a hyperlink and utilizing their e-mail address to login. Once logged in, participants had to read a consent form and click a check box indicating their consent to continue to the survey. Items were listed in separate rows with individual radio buttons for each response option. For the college student administration, the WDS was presented first, followed by the Iowa scales. Demographic information was collected last. [A screenshot of the survey is provided in Figure 2.]

For the professional administration, participants had 1 month to complete the survey, with a reminder e-mail sent out 2 weeks after the initial invitation. In


The Life Experience Self-Assessment

Please remember to submit your responses by clicking on the button at the bottom of the form.
Please rank the following statements from 1 = strongly disagree to 7 = strongly agree.

	strongly disagree . . . strongly agree						
	1	2	3	4	5	6	7
1. I accept constructive criticism	○	○	○	○	○	○	○
2. I accept others	○	○	○	○	○	○	○
3. I accept there are uncertainties in life	○	○	○	○	○	○	○
4. I achieve my goals	○	○	○	○	○	○	○

Figure 2. Screenshot of the survey.

the case of the college student administration, a reminder e-mail was sent 1 week after the initial e-mail, and the survey was closed after 2 weeks of data collection. The difference in time allotted to complete the surveys was due to the professional organization's request that we limit e-mail reminders to once every 2 weeks. On the other hand, the college students' finals period was shortly after the 2 weeks we allotted for survey submission, thus we felt it inappropriate to extend the data collection into this time. There were no missing data in either the professional or the college student sample.

Analyses

To begin, separate confirmatory factor analyses were performed on each administration of the WDS to determine the construct validity of scores within the two different populations. Then, both models were estimated independently but in the same run, providing an assessment of data-model fit indices across both samples simultaneously. Next, a model was run with all factor loadings constrained to be equal across samples and the overall fit of this model was compared to the previous using a scaled chi-square difference test (Kline, 2005; Satorra & Bentler, 1999) due to the robust maximum likelihood estimation used. A statistically non-significant result for this test would support weak measurement invariance (Meredith, 1993). Next, strong measurement invariance was assessed across these two samples by restraining the unstandardized factor loadings and intercepts to be equal (Kline, 2005) and examining the scaled chi-square difference test. While some recommend using chi-square difference tests to evaluate measurement invariance at this level, there is evidence that large sample sizes and non-normality can affect these tests, thus fit indices, specifically the Comparative Fit Index (CFI), were also examined to determine the feasibility of this model (Cheung & Rensvold, 2002). A difference of .01 or less between model CFIs was considered evidence of measurement invariance. Then we examined the latent means to test the hypothesis that the professional sample would have higher latent means than the college student sample, providing support for discriminant validity. All of the above analyses were performed using Mplus version 4.1 (Muthén & Muthén, 2006). Criterion validity was assessed by computing mean scores on each WDS and Iowa subscale and then performing Pearson correlations using SPSS version 14.

Reliability of the latent constructs was assessed using the maximal reliability measure Coefficient *H* (see Hancock & Mueller, 2001). This provides an indication of the degree to which the construct is captured by the information found within its measured indicators, and it was computed by hand. When evaluating factor models, Coefficient *H* is a better measure of reliability than Cronbach's α. Using SPSS, Cronbach's α was calculated for the entire instrument as a means of illustrating the internal consistency of the scores one could form using the factors' measured indicators, as this was used in the criterion validity analyses.

RESULTS

Descriptive statistics for the professional and college student samples, including item means, standard deviations, skewness, and kurtosis values, can be found in Appendices B and C, respectively.

Confirmatory Factor Analyses

Given the substantive theory behind the instrument, confirmatory factor analyses were used to test the construct validity of the WDS in both the professional and college student administrations. The CFAs were conducted using the variance-covariance matrices and allowing the factors to covary, but not the error variances. In both cases robust maximum likelihood estimation was used due to non-normality in some item univariate distributions. The chi-square test of fit will be reported but not used to determine data-model fit due to its sensitivity to sample size (Kline, 2005). Instead, we report the chi-square/df ratio, the CFI, the Tucker-Lewis coefficient (TLI) the standardized root mean-square residual (SRMR) and the root mean square error of approximation (RMSEA). Using current standards, evidence of good data-model fit is provided when the chi-square/df ratio is less than 2.0 (Kline, 2005), the TLI and the CFI greater than or equal to .96, the SRMR less than .09, or the RMSEA less than or equal to .06 (Hu & Bentler, 1999). Hu and Bentler (1999) recommend using the SRMR in combination with either the CFI or RMSEA to determine data-model fit, and we will follow this guideline. However, there is some evidence that the CFI and TLI fit indices may degrade when models, such as these, include a large number of variables (Kenny & McCoach, 2003). In addition, the variance extracted from each latent factor was reported as another indicator of construct validity.

Professional Administration

Model one was the CFA for the professional sample. Table 2 illustrates that the chi-square/df and comparative fit indices (TLI and CFI) do not reach common standards, but using the SRMR and RMSEA according to Hu and Bentler's (1999) criteria, there is evidence of good data-model fit. The chi-square value may be inflated due to sample size (Kline, 2005) and the TLI and CFI measures may not be good indicators of data-model fit given the large number of items in the WDS. In support of one of our hypotheses, evidence for the Willingness to Learn factor was found with these data. The Coefficient *H* for each factor, factor correlations, and variance extracted can be found in Table 3. All factors had a Coefficient *H* above .842 except for Willingness to Learn, which had a value of .702. These are excellent reliability values. Data from this sample had a Cronbach's α of .928, which is also very high. In terms of individual items, standardized factor loadings (see Appendix C) were all above .400 except for one item in the judgment factor (standardized loading = .361) and three of the eight reverse-coded items. Items

Table 2. Model Fit Information

Model	Robust χ^2, df	Robust χ^2/df	CFI	TLI	SRMR	RMSEA
1. Professional administration only	23179.149, 2974	7.79	.747	.738	.061	.051
2. College student administration only	6648.793, 2974	2.24	.685	.674	.081	.061
3. Both samples, no constraints	30468.178, 5948	5.12	.739	.729	.063	.052
4. Both samples, factor loadings constrained	30582.814, 6027	5.07	.738	.733	.074	.052
5. Both samples, factor loadings and intercepts constrained	31009.048, 6098	5.09	.735	.732	.074	.052

designed to load negatively on their respective factors did so. All *p*-values for the factor loadings were less than .01. The amount of variance extracted from each latent factor ranged from .26 to .44.

College Student Administration

Model two was the CFA for the college student sample (see Table 2). Using the same criteria as described previously, there is evidence of good data-model fit based upon the SRMR and the robust RMSEA. Here again, the Willingness to Learn factor was supported, whereas in Brown and Greene (2006) it was not. Table 4 shows Coefficient *H*, and factor correlations, and variance extracted values for this model. All Coefficient *H* values were above .844 except for Willingness to Learn which was .737. Data from this sample had a Cronbach's α of .930. These reliabilities are well above Cronbach and Shavelson's (2004) recommendation. In terms of standardized factor loadings (see Appendix C) for the items, all were above .400 except for one item in emotional management (.328), one in judgment (.298), one in life knowledge (.003), and one reverse-coded item in altruism (–.355). Items designed to load negatively on their respective factors did so. All *p*-values for the factor loadings were below .01 except for one loading in life knowledge (above) which was statistically non-significant. The variance extracted from each latent factor ranged from .29 to .44.

Table 3. Variance Extracted, Factor Correlations, and Coefficient *H* (on Diagonal) for Professional Sample

	WL	A	LK	LS	EM	J	L	SK
WL	.702	.932	.874	.783	.663	.985	.813	.589
A		.900	.762	.693	.563	.876	.758	.529
LK			.880	.555	.519	.864	.702	.547
LS				.868	.651	.708	.856	.609
EM					.842	.594	.708	.453
J						.867	.805	.599
L							.862	.637
SK								.868
Variance extracted	.26	.37	.37	.38	.35	.35	.34	.44

*All correlations $p < .01$.
**WL = Willingness to Learn; A = Altruism; LK = Life Knowledge; LS = Life Skills; EM = Emotional Management; J = Judgment; L = Leadership; SK = Self-Knowledge

CFA Summary

These findings provide support for the construct validity of scores from the WDS for two different samples: a professional organization and a different undergraduate institution. A majority of the individual item standardized factor loadings were high, however those falling below .400 may need to be rewritten or dropped from future administrations. Following Hu and Bentler's (1999) standards, there is acceptable model-data fit. In addition, the Willingness to Learn factor was supported in these analyses, whereas in Brown and Greene (2006) it was not. However, it should be noted that this factor correlated quite highly with both the Altruism and Judgment factors in both samples. It may be the case that the Willingness to Learn factor is still not measured in a way that makes it distinct from other constructs in the model. The next hypothesis concerned whether the factor loadings were invariant across the two samples.

Measurement Invariance

While data-model fit, as measured by CFA results, is important, if scores are going to be compared across samples then there must be measurement invariance (Meredith, 1993). First, a baseline model, here called model three (see Table 2), is run with both samples estimated and no cross-sample constraints. The first level

Table 4. Variance Extracted, Factor Correlations, and Coefficient *H* (on Diagonal) for College Student Sample

	WL	A	LK	LS	EM	J	L	SK
WL	.737	.881	.852	.828	.606	.958	.812	.594
A		.921	.741	.644	.356	.875	.733	.472
LK			.849	.513	.341	.854	.683	.518
LS				.886	.610	.669	.857	.574
EM					.844	.462	.605	.413
J						.882	.797	.563
L							.858	.580
SK								.852
Variance extracted	.29	.39	.31	.43	.32	.37	.33	.44

*All correlations $p < .01$.

**WL = Willingness to Learn; A = Altruism; LK = Life Knowledge; LS = Life Skills; EM = Emotional Management; J = Judgment; L = Leadership; SK = Self-Knowledge

of measurement invariance is called weak invariance, and involves constraining factor loadings across samples to be equal; factor correlations were not constrained and were hypothesized to differ. With this model, number four in Table 2, using Hu and Bentler's (1999) criteria, there was acceptable data-model fit based upon the SRMR and RMSEA. A formal test of the hypothesis of weak measurement invariance involves comparing two models, one with the factor loadings constrained across samples and one without these constraints. The scaled chi-square difference test between models four and three was statistically nonsignificant (scaled $\chi^2_{diff}(79) = 83.76741$, $p = .34$), suggesting that the more parsimonious model with constrained factor loadings be retained as a reasonable approximation of the relations among the data. In addition, the CFI value did not decrease by more than .01 (Cheung & Rensvold, 2002). Thus, we retained the hypothesis of weak measurement invariance of the factor loadings.

Model five imposed a means structure on the model, with constraints on factor loadings and intercepts across samples. A model with these constraints tests what Meredith (1993) calls strong measurement invariance. The scaled chi-square difference test between models five and four was statistically significant (scaled $\chi^2_{diff}(71) = 436.6205$, $p < .001$). However, model five did have acceptable fit using Hu and Bentler's (1999) criteria, and the CFI value increased by less than .01. Thus, following Cheung and Rensvold's (2002) guidelines, we retained

the hypothesis of measurement invariance of the factor loadings and intercepts and moved to interpreting the results of the latent factor means.

Discriminant Validity

We also hypothesized that participants in the professional sample, on average, would have higher scores on the eight latent factors than the college student sample. To compare latent means across samples, strong invariance is required, where both the factor loadings and intercepts must be constrained to be equal (Meredith, 1993), and this model was supported as shown previously. The professional sample was designated the comparison group, with latent means fixed to zero for identification purposes. As can be seen in Table 5, the latent means of the college student sample were all statistically significantly different from zero, and each was lower than those of the professional sample. These findings support our hypothesis that the professional sample's latent means would be higher than those of the college student sample, showing discriminant validity evidence for scores from the WDS.

Criterion Validity

Finally, we hypothesized that summed scores from three subscales of the WDS would correlate with scores from the Iowa Student Development Inventories subscales. The Cronbach's α reliabilities for the Iowa subscale for developing autonomy interdependence, developing purpose skills of life and managing emotions scales were .83, .84, and .89, respectively. Subscores were computed for

Table 5. Latent Means Model Results for College Student Sample

Latent factor	Latent mean	SE
Willingness to Learn	–.754	.094
Altruism	–.512	.077
Life Knowledge	–.558	.070
Life Skills	–.695	.080
Emotional Management	–.420	.071
Judgment	–.545	.079
Leadership	–.552	.074
Self-Knowledge	–.218	.074

Note: All latent means statistically significantly different from zero with $p < .01$.

the WDS and Iowa scales hypothesized to correlate. The Iowa subscale for developing autonomy interdependence was hypothesized to correlate with the altruism scale on the WDS and this correlation was statistically significant ($r = .56$, $p < .001$). The Iowa developing purpose scale and the WDS life skills scales correlated ($r = .51$, $p < .001$) as hypothesized. Finally, the Iowa managing emotions scale and the WDS emotional management scales also correlated ($r = .27$, $p < .001$) although this relation was smaller than the others. In sum, all of the hypothesized relations between the Iowa scales and the WDS were statistically significant, demonstrating evidence for the criterion validity of the WDS.

DISCUSSION

In this study we analyzed responses from professionals and college students to examine the construct validity and reliability of scores from Brown and Greene's (2006) WDS. This study extended previous work by utilizing multiple subpopulations and a large sample size to test whether the scores from the WDS could discriminate between groups and be used to produce valid inferences regarding wisdom as defined by Brown (2004b). Within the limitations of this study, discussed later, evidence of the construct validity and reliability of scores from the WDS with both the professional and collegiate sample was found. Acceptable data-model fit indices were found both individually and with the samples combined. The Willingness to Learn factor, not identified in a previous study (Brown & Greene, 2006), was supported here but its strong correlations with other subscales calls into question whether it has been sufficiently differentiated and described. The reliabilities of the subscale factors were good, but the variance extracted by each subscale varied from acceptable to low. Further evidence of the strength of the WDS was the finding of measurement invariance of factor loadings and intercepts across samples, although stronger evidence would be a statistically non-significant scale chi-square test for the models with and without fixed intercepts (models five and four, respectively, see Table 2). Latent means analyses supported the hypothesis that the professional sample would have statistically significantly higher latent factor scores than the college student sample, providing evidence of discriminant validity. Criterion validity evidence was also presented, as WDS subscales hypothesized to correlate with select Iowa Student Development Inventory subscales (Hood, 1997) did so. All of these findings provide support for the WDS as a measure of wisdom as defined by Brown's Model of Wisdom Development (Brown, 2004b).

Limitations

Certainly a major concern for both samples is the response rate. We were satisfied with the response rate for the professional sample, but not so with the collegiate sample. However, aside from a greater proportion of women in the collegiate sample than is found in the population, we believed that both samples

were representative, and therefore we proceeded with the analyses with this understanding (Cook et al., 2000). Another limitation of the study lies in the low variance extracted across subscales. While the factor loadings of each subscale were for the most part strong and above .400, item analysis may be necessary to increase the loadings in an effort to bolster the variance extracted. Likewise, the discriminant validity of the Willingness to Learn factor subscale remains in question. However, the professional sample was from an educational organization, and the college students obviously were currently pursuing learning, suggesting a possible lack of variance in responses to items measuring this factor.

Finally, we must acknowledge that the WDS has been tested solely with populations from a Western culture. There have been several cross-cultural examinations of wisdom (Takahashi, 2000; Takahashi & Overton, 2002, 2005), and their results suggest that across cultures older adults function at a higher level than their middle-aged counterparts. Additionally, these researchers argue that there are some specific effects of culture within each dimension of wisdom. Demonstrating cross-cultural applicability of the WDS is a direction for future research.

Implications

This study provides support for using the WDS as a measure of integrative, holistic learning. The WDS can help researchers do a number of things. First, it can help identify whether individuals are developing wisdom, and identify what sorts of intrapersonal factors and experiences affect it. Potential influences include such things as age, gender, and socioeconomic background. The WDS may assist in identifying what sorts of experiences seem to be more likely to promote the development of wisdom, in all aspects of human activity. These experiences can include work, school, relationships, community engagement, and religious and spiritual involvements. Ideally, this might inform societal and educational leaders, as well as others interested in human development, in developing more integrative and holistic learning experiences through their policies and programs, and provide greater capacity to assess them more pointedly. Such data may help substantiate decisions to allocate resources in our current empirically based decision-making environment.

CONCLUSION

As societal institutions seek to examine their influence upon their constituents' cognition, behavior, emotion, and morality, assessments of more integrated constructs, such as wisdom, are needed. Creating adequate psychometric measures of wisdom is a daunting but necessary task. This study contributes to a growing corpus of research (Brown, 2004b; Brown & Greene, 2006) suggesting that the Wisdom Development Scale has the psychometric qualities necessary to allow defendable inferences regarding participants' relative levels of wisdom based upon their performance on this measure. While psychometric investigations of

validity should never cease completely (Messick, 1989), based upon the corpus of research on the WDS, we feel applied research with the scale is warranted. Overall, it may be wise to utilize quantitative, self-report measures of wisdom in tandem with other types of measures such as observations and wisdom-based performance assessments (Baltes & Smith, 2008). Nonetheless, we have shown that the WDS can play an important role in a comprehensive evaluation of institutions and interventions designed to foster the development of wisdom.

APPENDIX A:
Sample Items from the Wisdom Development Scale

Self-Knowledge: I am well aware of my values.
Emotional Management: I get upset easily
Altruism: I show appreciation toward others
Leadership: I inspire others
Judgment: I take the context of the situation into consideration when making decisions
Life Knowledge: I look for deeper meaning of events in life
Life Skills: I handle multiple obligations effectively
Willingness to Learn: I seek assistance when necessary

APPENDIX B:
Descriptive Statistics and Factor Loadings for Professional Sample

Scale/Item	Mean	Standard deviation	Skewness	Kurtosis	Standardized factor loading
Willingness to Learn 1	5.66	.919	−1.083	2.671	.467
Willingness to Learn 2	2.17	1.204	1.541	2.769	−.397
Willingness to Learn 3	6.12	.987	−1.316	2.405	.487
Willingness to Learn 4	6.24	.768	−1.502	5.672	.695
Willingness to Learn 5	5.96	.953	−1.327	3.246	.542
Willingness to Learn 6	5.55	1.105	−.874	.980	.430
Altruism 1	6.15	.802	−1.327	5.791	.623
Altruism 2	2.32	1.260	1.648	2.950	−.317
Altruism 3	1.91	1.176	2.146	5.516	−.438
Altruism 4	5.92	.764	−1.054	3.625	.605
Altruism 5	6.21	.857	−1.562	4.801	.700
Altruism 6	6.21	.819	−1.615	6.008	.642
Altruism 7	6.30	.763	−1.636	6.384	.705
Altruism 8	5.43	1.055	−.813	.947	.533
Altruism 9	4.69	1.284	−.422	−.251	.454
Altruism 10	5.68	.931	−1.008	2.310	.616
Altruism 11	6.03	.880	−1.244	3.449	.681
Altruism 12	5.90	.938	−1.163	2.575	.617
Altruism 13	6.29	.739	−1.756	7.955	.750
Altruism 14	6.00	.840	−1.263	4.146	.664

Life Knowledge 1	6.26	.932	−1.749	4.900	.487
Life Knowledge 2	5.90	.975	−1.144	2.346	.600
Life Knowledge 3	5.48	1.033	−.615	.507	.539
Life Knowledge 4	5.82	.913	−1.006	2.505	.701
Life Knowledge 5	5.73	1.153	−1.022	1.182	.682
Life Knowledge 6	350	1.479	.332	−.739	−.185
Life Knowledge 7	5.77	1.114	−1.036	1.327	.680
Life Knowledge 8	6.00	.968	−1.303	2.978	.608
Life Knowledge 9	5.67	1.211	−.918	.635	.588
Life Knowledge 10	5.94	.917	−1.103	2.492	.746
Life Knowledge 11	5.83	1.060	−1.024	1.517	.702
Life Skills 1	5.63	.827	−.824	2.305	.591
Life Skills 2	5.55	1.026	−.832	.966	.542
Life Skills 3	2.76	1.453	.729	−.197	−.546
Life Skills 4	6.09	.837	−1.385	4.438	.618
Life Skills 5	5.78	.902	−1.061	2.573	.733
Life Skills 6	5.74	1.076	−1.095	1.791	.599
Life Skills 7	6.25	.905	−1.599	4.028	.620
Life Skills 8	5.66	.822	−1.086	2.736	.653
Life Skills 9	5.76	1.068	−1.148	1.933	.630
Life Skills 10	5.57	1.018	−.986	1.658	.669
Life Skills 11	5.60	1.001	−.766	1.115	.556
Emotional Management 1	4.84	1.244	−.424	−.180	.549
Emotional Management 2	5.02	1.359	−.596	−.082	.469
Emotional Management 3	5.79	.898	−.903	1.795	.545
Emotional Management 4	4.53	1.459	−.345	−.534	.510
Emotional Management 5	2.93	1.415	.734	−.046	−.588

APPENDIX B (Cont'd.)

Scale/Item	Mean	Standard deviation	Skewness	Kurtosis	Standardized factor loading
Emotional Management 6	3.46	1.503	.308	–.685	–.495
Emotional Management 7	5.45	1.096	–.800	.786	.621
Emotional Management 8	5.21	1.078	–.881	.945	.746
Emotional Management 9	4.97	1.181	–.765	.520	.727
Judgment 1	6.31	.855	–1.914	6.766	.531
Judgment 2	5.99	1.061	–1.445	3.144	.468
Judgment 3	2.09	1.199	2.009	4.786	–.374
Judgment 4	5.88	1.076	–1.965	5.677	.361
Judgment 5	5.92	.890	–1.157	3.118	.641
Judgment 6	5.79	.860	–1.027	2.900	.649
Judgment 7	6.03	.827	–1.424	4.991	.614
Judgment 8	6.07	.825	–1.351	4.759	.693
Judgment 9	6.38	.760	–1.985	8.503	.654
Judgment 10	6.36	.813	–2.052	8.485	.682
Judgment 11	6.52	.706	–2.679	14.852	.699
Leadership 1	5.61	.980	–1.146	2.968	.504
Leadership 2	5.75	.829	–.975	2.621	.605
Leadership 3	5.75	.880	–.992	2.547	.621
Leadership 4	2.61	1.408	1.151	.819	–.457
Leadership 5	5.60	.879	–.707	1.139	.538
Leadership 6	5.86	.869	–1.220	3.710	.638
Leadership 7	5.26	.971	–.527	.821	.570

Leadership 8	5.47	.927	−.623	.949	.634
Leadership 9	5.58	.975	−.874	1.488	.596
Leadership 10	5.90	.890	−1.012	2.354	.604
Leadership 11	5.71	.979	−.872	1.287	.621
Self-Knowledge 1	5.57	.974	−.798	1.293	.828
Self-Knowledge 2	5.81	.897	−1.038	2.343	.841
Self-Knowledge 3	5.28	1.080	−.664	.390	.556
Self-Knowledge 4	2.35	1.323	1.244	1.288	−.439
Self-Knowledge 5	5.34	1.065	−.769	.748	.602
Self-Knowledge 6	5.33	1.009	−.826	.912	.642

APPENDIX C:
Descriptive Statistics and Factor Loadings for College Student Sample

Scale/Item	Mean	Standard deviation	Skewness	Kurtosis	Standardized factor loading
Willingness to Learn 1	5.41	1.194	–.911	1.126	.561
Willingness to Learn 2	2.69	1.431	.757	–.024	–.447
Willingness to Learn 3	5.62	1.316	–1.105	1.210	.534
Willingness to Learn 4	5.95	.990	–1.219	2.270	.736
Willingness to Learn 5	5.36	1.341	–.935	.684	.411
Willingness to Learn 6	5.09	1.404	–.770	.194	.484
Altruism 1	5.92	1.069	-1.389	2.970	.593
Altruism 2	2.72	1.422	.957	.391	–.355
Altruism 3	2.28	1.367	1.399	1.586	–.533
Altruism 4	5.70	.986	–.756	.690	.614
Altruism 5	5.93	1.182	–1.598	3.205	.671
Altruism 6	5.85	1.099	–.927	.616	.675
Altruism 7	5.94	.997	–1.000	1.116	.664
Altruism 8	5.26	1.276	–.638	.310	.542
Altruism 9	4.27	1.599	–.128	–.938	.408
Altruism 10	5.43	1.135	–.792	.803	.608
Altruism 11	5.65	1.137	–1.113	1.899	.664
Altruism 12	5.77	1.007	–.889	1.116	.742
Altruism 13	6.02	.988	–1.614	4.664	.795
Altruism 14	5.58	1.113	–.870	1.063	.746

Life Knowledge 1	6.04	1.107	−1.448	2.261	.355
Life Knowledge 2	5.45	1.265	−.651	.186	.514
Life Knowledge 3	5.23	1.140	−.548	.480	.507
Life Knowledge 4	5.54	1.159	−.854	.684	.595
Life Knowledge 5	5.39	1.352	−.829	.339	.713
Life Knowledge 6	4.07	1.513	.096	−.538	.003
Life Knowledge 7	5.32	1.294	−.649	.027	.649
Life Knowledge 8	5.50	1.270	−1.044	1.312	.521
Life Knowledge 9	5.34	1.384	−.772	.259	.600
Life Knowledge 10	5.59	1.156	−.996	1.212	.693
Life Knowledge 11	5.29	1.275	−.383	−.588	.611
Life Skills 1	5.36	1.096	−.852	.949	.715
Life Skills 2	5.55	1.104	−.911	1.171	.571
Life Skills 3	3.85	1.772	−.076	−1.102	−.528
Life Skills 4	5.75	1.004	−.982	1.353	.633
Life Skills 5	5.34	1.128	−.663	.536	.736
Life Skills 6	5.39	1.332	−.854	.584	.603
Life Skills 7	5.47	1.367	−.967	.544	.667
Life Skills 8	5.29	1.054	−.649	.596	.680
Life Skills 9	5.08	1.334	−.547	.006	.626
Life Skills 10	4.95	1.404	−.566	−.269	.644
Life Skills 11	5.37	1.223	−.765	.453	.674
Emotional Management 1	4.78	1.466	−.402	−.468	.495
Emotional Management 2	4.77	1.580	−.452	−.559	.328
Emotional Management 3	5.56	1.059	−.584	.012	.494
Emotional Management 4	3.63	1.564	.218	−.870	−.491
Emotional Management 5	3.77	1.579	.053	−.811	−.479

APPENDIX C (Cont'd.)

Scale/Item	Mean	Standard deviation	Skewness	Kurtosis	Standardized factor loading
Emotional Management 6	4.23	1.629	–.294	–.694	.502
Emotional Management 7	5.06	1.412	–.649	–.171	.648
Emotional Management 8	4.90	1.324	–.5442	–.181	.739
Emotional Management 9	4.66	1.378	–.406	–.327	.799
Judgment 1	6.08	1.033	–1.252	1.826	.476
Judgment 2	5.66	1.195	–.906	.811	.547
Judgment 3	2.36	1.389	1.304	1.257	–.449
Judgment 4	5.76	1.242	–1.584	3.217	.298
Judgment 5	5.69	1.058	–.860	.803	.629
Judgment 6	5.50	1.015	–.700	.807	.693
Judgment 7	5.73	1.030	–.854	.804	.624
Judgment 8	5.74	.942	–.825	1.051	.705
Judgment 9	6.15	1.058	–1.721	3.681	.683
Judgment 10	6.03	1.171	–1.752	3.984	.693
Judgment 11	6.12	1.075	–1.850	4.528	.767
Leadership 1	5.28	1.177	–.787	.977	.543
Leadership 2	5.40	1.090	–.750	.669	.626
Leadership 3	5.60	.977	–.810	2.188	.558
Leadership 4	3.18	1.608	.657	–.543	–.398
Leadership 5	5.44	1.141	–.952	1.692	.516
Leadership 6	5.55	1.122	–1.205	2.114	.678
Leadership 7	4.87	1.210	–.487	.594	.564

Leadership 8	5.17	1.056	–.504	.037	.647
Leadership 9	5.46	1.134	–.894	1.081	.577
Leadership 10	5.55	1.051	–.563	–.026	.574
Leadership 11	5.05	1.298	–.521	.033	.598
Self-Knowledge 1	5.45	1.241	–.863	.776	.788
Self-Knowledge 2	5.71	1.123	–.995	1.023	.827
Self-Knowledge 3	5.18	1.181	–.551	.138	.574
Self-Knowledge 4	2.81	1.530	.760	–.105	–.459
Self-Knowledge 5	5.21	1.242	–.758	.486	.607
Self-Knowledge 6	5.03	1.182	–.710	.710	.629

ACKNOWLEDGMENT

The authors would like to thank Dr. Roger Azevedo for his support of this project.

REFERENCES

Ardelt, M. (1997). Wisdom and life satisfaction in old age. *Journal of Gerontology, 52,* 15-27.

Ardelt, M. (2003). Empirical assessment of a three-dimensional scale. *Research on Aging, 25,* 275-324.

Ardelt, M. (2004). Wisdom as expert knowledge system: A critical review of a contemporary operationalization of an ancient concept. *Human Development, 47,* 257-285.

Baltes, P. B., & Smith, J. (1990). Toward a psychology of wisdom and its ontogenesis. In R. Sternberg (Ed.), *Wisdom: Its nature, origins and development* (pp. 87-120). Cambridge: Cambridge University Press.

Baltes, P. B., & Smith, J. (2008). The fascination of wisdom: Its nature, ontogeny, and function. *Perspectives on Psychological Science, 3*(1), 56-64.

Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American Psychologist, 55,* 122-136.

Brown, S. C. (2002a, Fall). Developing students' multiple identities: How career services practitioners can put theory into practice. *Journal of College and Employers,* 28-32.

Brown, S. C. (2002b, Summer). A model for wisdom development and its place in career services. *Journal of College and Employers,* 29-36.

Brown, S. C. (2004a). Where this path may lead: Understanding post-college decision-making. *Journal of College Student Development, 45,* 375-390.

Brown, S. C. (2004b). Learning across campus: How college facilitates the development of wisdom. *Journal of College Student Development, 45,* 134-148.

Brown, S. C. (Winter, 2006). Educator or entrepreneur? Marketing and other strategies to increase career learning outcomes. *NACE Journal, 66*(2)*,* 26-34.

Brown, S. C., & Greene, J. A. (2006). The wisdom development scale: Translating the conceptual to the concrete. *Journal of College Student Development, 47,* 1-19.

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

Chickering, A. W., & Reisser, L. (1993). *Education and identity* (2nd ed.). San Francisco: Jossey-Bass.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement, 60,* 821-836.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64,* 391-418.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed., Vol. 26). Thousand Oaks, CA: Sage.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A festschrift in honor of Karl Jöreskog* (pp. 195-216). Lincolnwood, IL: Scientific Software International, Inc.

Hayduk, L. A., & Glaser, D. N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling, 7*, 1-35.

Holliday, S. G., & Chandler, M. J. (1986). *Wisdom: Explorations in adult competence.* New York: Karger.

Hood, A. B. (1997). *The Iowa student development inventories* (2nd ed.). Iowa City, IA: Hitech Press.

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

Kenny, D. A., & McCoach, B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*(3), 333-351.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Kunzmann, U., & Baltes, P. B. (2003). Wisdom-related knowledge: Affective, motivational, and interpersonal correlates. *Personality and Social Psychology Bulletin, 29,* 1104-1119.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525-543.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Muthén, B. O., & Muthén, L. K. (2006). *Mplus Version 4.1*. Los Angeles, CA: Muthén & Muthén.

National Research Council, Committee on Scientific Principles for Education Research. (2002). *Scientific research in education.* Washington, DC: National Academy Press.

Satorra, A., & Bentler, P. M. (1999). A scaled difference chi-square test statistic for moment structure analysis. *UCLA Statistics Series #2,* University of California, Los Angeles.

Schuman, D. (1982). *Policy analysis, education, and everyday life: An empirical reevaluation of higher education in America.* Lexington, MA: D. C. Heath and Company.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55,* 5-14.

Smith, J., & Baltes, P. B. (1990). Wisdom-related knowledge: Age/cohort differences in response to life-planning problems. *Developmental Psychology, 26,* 494-505.

Staudinger, U. M., Maciel, A. G., Smith, J., & Baltes, P. B. (1998). What predicts wisdom-related performance? A first look at personality, intelligence and facilitative experiential contexts. *European Journal of Personality, 12,* 1-17.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49,* 607-662.

Sternberg, R. J. (1990). Wisdom and its relations to intelligence and creativity. In R. J. Sternberg (Ed.), *Wisdom: Its nature, origins and development* (pp. 142-159). Cambridge: Cambridge University Press.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized.* New York: Cambridge University Press.

Takahashi, M. (2000). Toward a culturally inclusive understanding of wisdom: Historical roots in the East and West. *International Journal of Aging and Human Development, 51*(3), 217-230.

Takahashi, M., & Overton, W. F. (2002). Wisdom: A culturally inclusive developmental perspective. *International Journal of Behavioral Development, 26*, 269-277.

Takahashi, M., & Overton, W. F. (2005). Cultural foundations of wisdom. In R. J. Sternberg & J. Jordan (Eds.), *A handbook of wisdom: Psychological perspectives* (pp. 32-60). New York: Cambridge University Press.

Webster, J. D. (2003). An exploratory analysis of a self-assessed wisdom scale. *Journal of Adult Development, 10,* 13-22.

Webster, J. D. (2007). Measuring the character strength of wisdom. *International Journal of Aging and Human Development, 65*, 163-183.

Wink, P., & Helson, R. (1997). Practical and transcendent wisdom: Their nature and some longitudinal findings. *Journal of Adult Development, 4,* 1-15.

Direct reprint requests to:

Jeff Greene
School of Education
CB#3500
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3500
e-mail: jagreene@email.unc.edu