

MULTIMETHOD APPROACH FOR MEASURING CHANGES IN CHARACTER

William H. Hendrix and Christopher J. Luedtke

United States Air Force Academy

Cassie B. Barlow

Air Force Senior Leader Management Office

This research involved 3 studies performed to develop and validate 2 instruments for measuring moral excellence or moral character: Behavioral Desirability Scale (BDS) and the Character Assessment Rating Scale (CARS). The first study dealt with instrument development; the second focused on instrument validation; and the third provided a replication of the validation study in a different setting with a different population.

A recent survey in the United States focused on the need for and importance of moral education. The survey was conducted by *Who's Who Among American High School Students* (Kleiner & Lord, 1999), and the results indicated that 80% of high-achieving high school students admitted that they cheated at least once. Half of these students indicated that they did not feel cheating was wrong. In addition, Kleiner & Lord (1999) noted that research from the *Who's Who Among American High School Students* survey reveals that 75% to 98% of college students admitted to cheating in college.

These studies stress the need for developing moral character or moral excellence within our youth. The major questions are how do we

develop the character of students and how do we know we have been successful? What are the methods we should use and how do we measure changes in character? Central to developing character is incorporating within character education programs the components that have been shown to work. Berkowitz (2002) provides a very useful summary of what appears to be the most promising interventions for developing individuals' character (e.g., guided peer moral discussions, service learning, building a positive community). He notes that there is little agreement about how to define character and what components make up character. Berkowitz (2002) defines character as "an individual's set of psychological characteristics that affect that person's ability

• William H. Hendrix, 18270 Knowlwood Blvd., Monument, CO 80132, Phone: 719-481-9888.

Journal of Research in Character Education, 2(1), 2004, pp. 59–80
Copyright © 2004 Information Age Publishing, Inc.

ISSN 1543-1223
All rights of reproduction in any form reserved.

and inclination to function morally.” He addresses what it means to be a moral person in terms of what he calls a moral anatomy, which is made up of “moral behavior, moral values, moral personality, moral emotion, moral reasoning, moral identity, and foundational characteristics.” Similarly, Lickona (1991) has suggested a model made up of cognition, affect, and behavior or moral knowing, moral feeling, and moral behavior. These conceptions of character or moral excellence provide a useful framework for character development programs.

Pascarella and Terenzini (1991) reviewed the empirical studies that examined the influence of college on students’ attitudes. Their work shows that the college experience has significant impact on college students’ attitudes, values, and behaviors. They found a consistent pattern across research studies. Furthermore, they noted that research consistently shows changes during college of increased humanitarianism altruism, sense of social conscience and civic responsibility. Similarly, Astin (1993) reviewed research on the affects of college on student values and concluded that college had a profound impact on student attitudes and values. These included increased racial understanding, commitment to programs to clean up the environment, and greater commitment to developing a meaningful philosophy of life. Myrsky and Helkama (2002) investigated the effects of a professional ethics course on social psychology students’ moral sensitivity when presented with a series of moral issues. They found significant increases in moral sensitivity for the students in the course compared to a control group. These research results are consistent with that proposed by Berkowitz (2002) and Lickona (1991) that values associated with moral behavior can be influenced by education.

As pointed out by Walker and Pitts (1998), contemporary moral psychology models (e.g., Eisenberg, 1995; Gilligan, 1982; Kohlberg, 1981, 1984; Shweder, Mahapatra, & Miller, 1987) have focused more on moral reasoning and have paid little attention to the construct of

moral excellence. Moral excellence or moral character is broader conceptually than the construct of moral reasoning. The construct of moral excellence incorporates moral cognition and also deals with the affective and behavioral domains. Berkowitz’s (2002) moral anatomy concept and Lickona’s (1991) tripartite model reflect the broader concept of character or moral excellence.

The need for character development is not limited to children and adolescents. The business world deals with critical issues that involve ethical decision making and moral behaviors. Ajzen and Fishbein (1980) and Ajzen (1985, 1991) proposed and expanded a theory of planned behavior to better explain ethical decision making. Key to this theory is the concept of intentions as a predictor of behavior. Intentions, they suggest, are shaped by the influence of important others and by a person’s attitudes toward ethical behaviors. Based on this theory, Flannery and May (2000) conducted research within a U.S. metal-finishing plant to determine what factors contribute to ethical decision making for environmental issues. Their results suggested a path-analytic model indicating that the magnitude of consequences for others moderated the relationship of five antecedents (ethical attitudes, subjective norms, self-efficacy, financial cost, and ethical climate) on manager’s ethical decision-making intentions. Therefore, intentions are useful in predicting ethical decision making and behaviors.

Another research need deals with character assessment. Specifically, a need exists for instruments and methods designed to measure individuals’ character or moral excellence in terms of cognitive, affective, and behavioral measures. To date, the array of character-related assessment instruments is very limited. Among those instruments available, most deal with moral reasoning (e.g., Defining Issues Test, DIT, and Moral Judgment Interview, MJII). A few have constructed instruments that attempt to measure all aspects of moral functioning including the student character questionnaires created by Vessels (1998)

and the “global portraits” developed by Davidson and Khmelkov (2003).

Leming (1993) noted the need for a “grand theory” of character education and research to guide future efforts. He also indicated that today the thought and research in this area is atheoretical which hampers the process of developing effective programs. There are a number of theoretical approaches and models that have been proposed to guide character education. These lead educators to focus on how to conduct programs that will possibly increase the development of moral character or moral excellence. Implied in these efforts is that there are measurable outcomes that can be assessed to measure the success of the educational process. It would appear reasonable that any development of assessment instruments should be based on the outcomes suggested by character education theory and literature. Battistich (1999) noted that most research on character development has focused narrowly on a single outcome such as moral reasoning or an undesirable outcome like cheating. Battistich contends that a better job must be done in representing the domain of moral character if we are to make progress in moral development and research. He asks the question: “What are the range of outcome variables that should be assessed?” In the spirit of this question, this research drew from a series of character-related theoretical models and literature to establish character-related dimensions suggested by them and that were frequently used in various character educational programs. Battistich (1999) suggested that the work of Rest (1983) and Berkowitz (1995) would be a reasonable starting point in developing a range of outcomes supported by theory. The fourfold typology of characteristics that Rest (1983) proposed lead to moral behavior includes moral sensitivity, moral values, moral judgment, moral motivation, and moral character. Berkowitz (1995) provided seven characteristics in describing the anatomy of a moral person. These include moral behaviors, moral values, moral character, moral reason, moral emotion, moral identity, and meta-moral char-

acteristics (i.e., ones that support moral functioning). In addition to these theoretical models, the “Just Community” model suggested by Kohlberg and his colleagues (e.g., Kohlberg & Higgins, 1987) provides a basis for establishing relevant outcome measures. Review of “Just Community” concepts include moral autonomy, moral maturity, respect for others, and altruism or caring for others. In a similar vein are the concepts underlying the theoretical model of the Child Development Project (Battistich, Solomon, Watson, & Schaps, 1997). This project focuses on caring communities of learners with program effectiveness designed to change students’ attitudes, motives, and behaviors. These include values of fairness, responsibility, and concern of the welfare of others. The above theoretical models, the theoretical model of Ajzen and Fishbein (1980) and Ajzen (1985, 1991) on ethical decision making, and character development literature served as a basis for selecting the outcome dimensions used in this research. By design, measurement of moral reasoning was not a part of this research, as reliable measures already exist (e.g., DIT, MJI). Instead the focus was on developing instruments to measure the broader concept of moral excellence or character.

Establishing the validity and reliability of character assessment instruments is imperative and challenging. The validity and reliability of the DIT has been studied for 20 years, and this research provides useful models for validation of other character assessment instruments. Results of the DIT validation research can be used to validate other character-related instruments and serve as a framework for the Behavioral Desirability Scale (BDS) and the Character Assessment Rating Scale (CARS).

Rest, Narvaez, Bebeau, and Thoma (1999) used six criteria for establishing construct validity for the DIT. These included (a) differentiation of groups differing in expertise, (b) longitudinal upward trends, (c) sensitivity to moral education interventions, (d) developmental hierarchy (higher is better), and (e) links to behavior, and (f) links to political atti-

tudes and political choice. As a group these criteria can be classified into two major types of validity: differential and convergent-divergent. Differential distinguishes between groups that logically should differ, and convergent shows that DIT moral reasoning scores correlate with other constructs related to moral reasoning.

Differential and convergent-divergent validity provide a useful framework for validation of the instruments developed as a part of this research. Of particular interest are the research results of Rest and Narvaez (1994) and Rest et al. (1999) who reported differential validity for age, education, and gender. Specifically, education was the most predictive factor associated with moral reasoning. An increase in formal education was associated with an increase in moral reasoning scores. Age was generally related to increased moral reasoning, but moral-reasoning scores showed an increase only as long as individuals were in formal education programs. Gender had the least effect with female scoring slightly higher than males. More recently, Bebeau (2002) reported a larger difference between males and females with females scoring higher on the DIT 2 than males.

Additional support for educational level being associated with increased moral reasoning comes from investigations using the Moral Judgment Interview (MJI) and the Moral Judgment Test (MJT). This research found a consistent positive relationship between educational level and moral reasoning in cross-sectional (Eiferman, 1982; Lei, 1981; Lei & Cheng, 1984; Mentkowski & Strait, 1983; Whitla, 1978) and longitudinal studies (Colby, Kohlberg, Gibbs, & Lieberman, 1983; Lind, 1985, 1986; Loxley & Whiteley, 1986; Mentkowski & Strait, 1983).

Rest and Narvaez (1994) also provide evidence for convergent and divergent validity of the DIT. They found the DIT's measure of moral reasoning (P-Score) was positively correlated with other measures of moral reasoning such as the Moral Judgment Interview and the Comprehension of Moral Concepts test. The

DIT's P-Score was not correlated with social desirability or with most personality trait measures (divergent validity). Finally, Rest and Narvaez (1994) and Rest et al. (1999) present evidence of concurrent and predictive validity. The P-Score has been found to be significantly correlated with a wide array of ethical and prosocial behaviors including community involvement and civic responsibility.

The reliability and validity of self-report instruments have also been examined. It is typically the case that some individuals taking self-report instruments such as surveys and questionnaires introduce measurement by not taking the process seriously. This includes random responding, marking in a pattern, or answering in a socially desirable manner. All of these activities introduce measurement error and need to be addressed. As reported in Rest et al. (1999) checks for marking responses in a pattern, or randomly, can be established. For social desirability the Crown and Marlowe (1960) social desirability scale can be added to instruments in an effort to determine if individuals are acting in a socially desirable manner or are being truthful. These processes used in establishing the validity and reliability of the DIT and the results of research using the MJI and MJT can serve as a guide for future instrument development including those described in this research.

This research grew out of a need at the United States Air Force Academy (USAFA) to develop a way to measure changes in the character of cadets. The measurement devices needed to be applicable for a wide variety of comparison settings, individuals, and groups. The research plan called for comparisons to be made between USAFA cadets and civilian university students to see if character-related changes differ between institutions over time. The study included making comparisons at the freshman and senior levels, and 3 years after graduation using Air Force active duty personnel, and USAFA cadets. The instruments developed and validated as a part of this research used the models proposed by Berkowitz (2002) and Lickona (1991) and the

concept of moral excellence proposed by Walker and Pitts (1998). These instruments, the Behavioral Desirability Scale and the Character Assessment Rating Scale, were designed to measure the cognitive and behavioral domains. The Behavioral Desirability Scale was designed to measure one aspect of the cognitive domain, that of individual character-related values. This instrument used a desirability scale ranging from 1 = Extremely Undesirable to 9 = Extremely Desirable. The CARS was designed to measure one aspect of the behavioral domain, that of character-related behaviors of individuals. There were two versions of the Character Assessment Rating Scale. One version was a self-rating version for use by the ratee while the other version was for ratings by other raters in a 360 degree format (e.g., supervisor, peer, and subordinate raters). The CARS asked individuals to rate a person on the frequency that an individual displays each of 12 dimensions of behavior with scale values ranging from 1 = Never to 8 = Always.

STUDY 1: INSTRUMENT DEVELOPMENT

Method

Participants

The participants in this study were 682 USAFA junior and senior level cadets. Data collected with this sample were used primarily for developing the Behavioral Desirability Scale (BDS) and the Character Assessment Rating Scale (CARS). For this sample approximately 82.9% were males and 17.1% females with an age range of 18 to 25 and an average age of 20.18 years. The racial composition was: American Indian or Alaskan Native, 2.3%; Asian or Pacific Islander, 5.3%; Black, not of Hispanic Origin, 6.6%; Hispanic, 5.9%; White, 75.1%; and those listing themselves as Other or not reporting their race, 4.8%.

Character Dimensions

An extensive review of the character-related literature was performed to identify the most frequently cited dimensions of character. This included journal articles, texts, and online sites. The character dimensions selected had to be consistently cited as important for the military and commonly cited in character-related literature. Based on this review 13 of the most frequently cited character dimensions were selected as the basis for instrument development. These dimensions are provided in Table 1 along with definitions. Two of these dimensions, Responsibility and Self-Discipline, were combined since they seemed to be logically related.

Procedure

Data were collected using paper copies and online versions for the Behavioral Desirability Scale (BDS) and the Character Assessment Rating Scale (CARS). Data were collected from volunteers en masse when using the paper copies of the instruments. For those using the online survey, participants were briefed on the project and provided with instructions. For both data collection methods, participants were provided with an informed consent document. For paper and online versions, a comment section was provided. Feedback was provided to participants in an article summarizing the results.

Survey Development

Behavioral Desirability Scale. The Behavioral Desirability Scale was designed to measure one aspect of the cognitive domain, that of individual character-related values. Students in four sections of a senior level psychology course were provided with the dimensions listed in Table 1. This was a class project that supported a topic on survey development and validation. Students were divided into three teams in each class section. Each team was given one third of the dimensions in Table 1

TABLE 1
Character Dimensions and Definitions

<i>Dimension</i>	<i>Definition</i>
Integrity	Consistently adhering to a moral or ethical code or standard. A person who consistently chooses to do the "right thing" when faced with alternate choices.
Honesty	Consistently being truthful with others.
Loyalty	Being devoted and committed to one's organization, supervisors, coworkers, and subordinates.
Selflessness	Genuinely concerned about the welfare of others and willing to sacrifice one's personal interest for others and their organization.
Compassion	Concern for the suffering or welfare of others and provides aid or shows mercy for others.
Competency	Capable of performing tasks assigned in a superior fashion and excels in all task assignments. Is effective and efficient.
Respectfulness	Shows esteem for and consideration and appreciation of other people.
Fairness	Treats people in an equitable, impartial, and just manner.
Responsibility and Self-discipline	Can be depended on to make rational and logical decisions and to do tasks assigned. Can perform tasks assigned without supervision.
Decisiveness	Capable of making logical and effective decisions in a timely manner. Does not "shoot from the hip," but does promptly make a good decision after considering data appropriate to the decision.
Spiritual Appreciation	Values the spiritual diversity among individuals with different backgrounds and cultures and respects all individuals' rights to differ from others in their beliefs.
Cooperativeness	Willingness to work or act together with others in accomplishing a task or some common end or purpose.

and asked to break each dimension down into subdimensions or subfactors. For example, responsibility might be broken down into financial responsibility, responsibility to family, responsibility to one's career, and responsibility to one's community. The resulting subdimensions were then grouped into similar categories, and a list was formed for each team's set of dimensions. In each class section, these subdimension lists were given back to the same team that had created them. Each team was then asked to write positive or desirable and negative or undesirable statements for each of the subdimensions. These were sorted into groups by similarity. Each of the resulting item topics was rewritten to be in a common style and placed into the first version of the BDS.

This version of the BDS contained 102 items and was administered to 682 volunteers in the 1999 entering class. Data from this administration were factor analyzed using principal-components factor analysis with varimax rotation to identify the underlying

character-related dimensions. This analysis produced four factors (Selflessness, Integrity, Competency, and Spiritual Appreciation). Selflessness measures a series of related concepts including selflessness, respectfulness, compassion for others, loyalty, and cooperativeness. Integrity measures an individual's integrity in terms of adhering to a set of moral or ethical standards and being consistently honest in dealing with others. Competency measures an individual's conscientiousness and competence in performing tasks assigned, performing all tasks effectively and efficiently, being responsible, making sound decisions, and treating others fairly. Spiritual Appreciation measures the extent that an individual values the spiritual differences of others that come from different backgrounds and cultures. Items with low factor loadings were deleted from the survey resulting in 65 remaining items. These 65 items and their factor loadings, from the 102-item BDS survey analysis, are provided in the Appendix along with the alpha reliability coefficients.

Data accumulated after the initial four-factor model development ($n = 5,729$) was later used to perform confirmatory factor analysis (Arbuckle & Wothke, 1999) for the BDS. The sample included high school students ($n = 152$), Air Force Academy freshmen ($n = 2,323$), Air Force Academy seniors ($n = 1,028$), Air Force Academy Preparatory School cadets ($n = 3,410$), US Air Force officers ($n = 1,395$), and U.S. Air Force noncommissioned officers ($n = 490$). Of these, 79.8% were males and 20.2% females. Participants were volunteers who had time set aside for them to complete the surveys en masse using paper versions of the survey. Confirmatory factor analysis was performed by imposing the factor model described above and then establishing the goodness-of-fit of the larger data sample to the model. The resulting analysis confirmed that the larger data set had a good fit to the original four-factor model giving additional support for the factor structure reported above. Specifically, the goodness-of-fit indexes for the BDS four factor model were: normed fit index = .97, relative fit index = .96, incremental fit index = .97, Tucker-Lewis index = .96, comparative fit index = .97.

Instructions for completing the BDS ask individuals to indicate how desirable or undesirable they consider each of the BDS items. The BDS consisted of a nine-point rating scale (1 = Extremely Undesirable, 5 = neither Desirable nor Undesirable, and 9 = Extremely Desirable). The logic associated with this approach is that those taking the BDS will communicate the value they place on each item. For example, consider an item that deals with a person finding and returning another persons property. A person responding to this item, whose value system would be considered of higher character, should see this event as more desirable, and rate this item higher or more desirable, than a person with a lower character-related value system.

Character Assessment Rating Scale. The CARS was designed to measure one aspect of the behavioral domain, that of charac-

ter-related behaviors of individuals. As initially conceived, the CARS assessment process was to be based on a 360 degree rating scheme whereby individuals rate themselves and are rated by others (superior, peer, and subordinate raters). A Behavioral Observation Rating Scale (BORS) format was used for the CARS. The BORS is a type of measurement scale used in personnel performance evaluation. This scale asks individuals to rate a person on the frequency that an individual displays certain character-related behaviors (e.g. integrity, honesty, etc.) with scale values generally ranging from 1 = Never to 8 = Always.

Instructions for completing the CARS asked individuals to rate themselves (Self-rating Form) or another person (Supervisor, Peer, and Subordinate Rating Form) on the 12 CARS dimensions. In the cadet system there is a chain of command similar to business organizations. Therefore, each cadet has a superior or supervisor, peers, and subordinates. Due to the small sample size of ratings by other raters (e.g., $n = 33$ full cases of supervisor, peer, and subordinate raters) this study was limited to the CARS self-ratings only.

The CARS Self-form data ($n = 682$) from the 1999 entering cadet class was also factor analyzed resulting in a single factor. Participants included Air Force Academy freshmen ($n = 568$) and seniors ($n = 114$). Approximately 85.8% were males and 14.2% females. Data ($n = 4,609$) collected after the original data collection was subjected to confirmatory factor analysis to see if the larger database fit the single factor model originally established. The sample included high school students ($n = 149$), Air Force Academy freshmen ($n = 1,665$), Air Force Academy seniors ($n = 868$), Air Force Academy Preparatory School cadets ($n = 332$), U.S. Air Force officers ($n = 1,130$), and U.S. Air Force noncommissioned officers ($n = 465$). Of these, 79.5% were males and 20.5% females. Participants were volunteers who had time set aside for them to complete the survey en masse using paper versions of the survey. The results indicated a very good

fit of the data to the single factor model (normed fit index = .99, relative fit index = .98, incremental fit index = .99, Tucker-Lewis index = .99, comparative fit index = .99).

Summary

This study described the process used in developing two surveys for assessing dimensions of character. The BDS was designed to measure cognitive or character-related values while the CARS was designed to measure rated behaviors. The next study describes the process used to validate these instruments.

STUDY 2: INSTRUMENT VALIDATION

The purpose of this study was to establish the validity and reliability of the two character assessment instruments. Validation included discriminant, convergent, and divergent validity. Establishing the reliability of the Behavioral Desirability Scale consisted of internal consistency reliability using Cronbach's Coefficient Alpha analysis.

Results of DIT research (Rest & Narvaez, 1994; Rest et al., 1999) has reported differential validity for education, age, and gender. Specifically, scores in moral reasoning increase with age, educational level, and gender with females scoring slightly higher than males. Therefore, the character assessment instruments used in this research should logically discriminate in a similar fashion. Age was not evaluated since age and educational level were highly correlated ($r = .95$) in this research sample. In addition, the USAF Academy focuses on character development during the 4 years cadets are attending the institution. The honor code is the basic foundation, but there are numerous programs of instruction, seminars, and conferences that are designed to build leaders of character. Therefore, it would be expected that senior cadets should score higher on character measures than entering freshmen. A select group of volunteer cadets are chosen to be honor representatives. These cadets effectively run the

honor system by providing honor training and education for other cadets. It would appear reasonable that these cadets should value the BDS factor of integrity more than cadets in general due to their desire to be an honor representative and being an active in the process.

Construct validity in terms of convergent and divergent validity involves correlating the overall BDS score with the CARS score. The following hypotheses were tested as a part of the validation process to establish discriminant validity:

Hypothesis 1: Scores on the BDS and CARS will increase with educational level. BDS scores will increase between freshmen cadets, senior cadets, and graduates (i.e., cadets that have been out of school for three years).

Hypothesis 2: Females will have higher scores than males on the BDS sub-scales, overall BDS Index, and the CARS Index.

Hypothesis 3: Honor representatives will score higher on the BDS subscale of integrity than cadets in general.

Hypothesis 4: The BDS and the CARS will not be subject to participants responding in a socially desirable manner.

The following hypotheses were tested as a part of the validation process to establish convergent and divergent validity:

Hypothesis H5: The BDS Index will be correlated positively with the CARS Index.

Hypothesis H6: The BDS and CARS Indices will be positively correlated with the moral reasoning indices of the DIT.

Method

Participants

The number of participants completing the two character assessment instruments varied for each instrument. This was due to two of the instruments (Behavioral Desirability Scale and Character Assessment Rating Scale) being

available in paper form as well as online. Some groups of participants were available in large rooms and did not have access to computers or they had not yet been provided with access to computers linked to the Academy's intranet. Participants used in validation and other analyses consisted of 4,702 individuals for the Behavioral Desirability Scale and 3,633 for the Character Assessment Rating Scale (Self-form). Participants for the Behavioral Desirability Scale included Air Force Academy freshmen ($n = 2,323$), Air Force Academy seniors ($n = 643$), Air Force Academy Preparatory School cadets ($n = 341$), and U.S. Air Force officers ($n = 1,395$). Participants were volunteers who had time set aside for them to complete the survey en masse using paper versions of the survey. The character assessment rating scale consisted of the same group of participants. The age for cadets ranged from 17 to 25 years with a mean age of 20.3 years. Data on active duty officers were obtained on those in professional military education programs. The average age of these active duty officers ranged from approximately 26 to 47 years based on class group averages. Approximately 76.9% of the participants were males while 20.5% were females and 2.6% did not provide their gender. The racial composition was: American Indian or Alaskan Native, 1.7%; Asian or Pacific Islander, 5.5%; Black, not of Hispanic Origin, 6.6%; Hispanic, 7.6%; White, 73.9%; and those listing themselves as Other or not reporting their race, 4.4%.

Procedure

The procedure used for collecting data for this study was the same as in Study 1. Data were collected using paper copies and online versions for the Behavioral Desirability Scale (BDS) and the Character Assessment Rating Scale (CARS). Data were collected from volunteers en masse when using the paper copies of the instruments. For those using the online survey, participants were briefed on the project and provided with instructions. For both data collection methods, participants were provided

with an informed consent document. For paper and online versions, a comment section was provided. Feedback was provided to participants in an article summarizing the results.

Discriminant Validation

Analyses to establish discriminant validity involved two different categories of participants: Air Force officers and USAFA cadets. The first category consisted of Air Force officers attending education and training programs at the Air Force's Air University. This sample was selected since it would provide a wide range of educational levels as opposed to cadets who had less than a bachelor's degree. In addition, personnel in this group came from a variety of colleges and universities and, therefore, were more diverse. This group's educational level ranged from those who had recently graduated from college to those with advanced degrees. Four educational groups were established for analysis. These included: (a) bachelor's degree, (b) bachelor's degree plus advanced professional training, (c) master's degree level, and (d) post-graduate education plus extensive managerial and leadership education. The first group (bachelor's degree) consisted of recent college graduates who graduated within 18 months of serving as participants. The second group (bachelor's degree plus advanced professional training) had graduated from college 5 to 8 years earlier and had received advanced professional training related to management and leadership. The third group (master's degree) generally had obtained a master's degree at a civilian university but also were attending a program generally considered to be equivalent to a master's degree. The last group consisted of senior officers (i.e., brigadier generals) who had advanced education beyond that of a master's level degree.

The second category of participants included USAFA cadets at different levels in their educational development and USAFA graduates. New cadets (entering freshmen), senior cadets nearing graduation, and graduates of the Air Force Academy 3 years after

graduation participated in the research. These three groups were surveyed in order to see if the survey instruments were able to discriminate between them.

Lastly, we would expect that cadets serving as honor representatives would value the BDS factor of integrity more than cadets in general.

Results

The first hypothesis, *scores on the BDS and CARS will increase with educational level*, was tested for differences between groups. This analysis was an extension of research performed by Barlow, Jordan, and Hendrix (2003) that deals with leadership and character. The response options for the BDS range from 1 (extremely undesirable) to 9 (extremely desirable). The overall BDS scale and the BDS subscales (i.e., selflessness, integrity, competency, spiritual appreciation), however, was

effectively a five-point scale ranging from 5 to 9 since negatively worded items were reversed scored and an average of all items were computed for each subscale and overall BDS scale. The mean of the overall BDS scale ($n = 5,087$) was 7.29 with 95% of responses falling between 5.89 and 8.69. For the CARS Self-form, the response options ranged from 1 (Never) to 8 (Always). The mean of the CARS Self-form scale was 5.33 ($SD = 1.26$). The 360-degree CARS index was not used in the analysis due to a very small sample ($n = 33$). Even though over 700 CARS surveys were given out to other raters to rate a focal person, still only 33 full cases (a superior rater, one or more peer raters, and one or more subordinate raters) were received. Nevertheless, the inter-correlation among those received and with the CARS Self-form ranged from .83 to .90 indicating that other raters' scores were similar

TABLE 2
BDS and CARS Differences by Educational Level

<i>Factor</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Selflessness				
Between Groups	33.57	3	11.19	15.57
Within Groups	745.09	1,037	.72	
Total	778.66	1,040		
Integrity				
Between Groups	106.26	3	35.42	40.57
Within Groups	905.43	1,037	.87	
Total	1,011.69	1,040		
Competency				
Between Groups	30.30	3	10.10	14.80
Within Groups	707.36	1,037	.68	
Total	737.65	1,040		
Spiritual Appreciation				
Between Groups	17.69	3	5.90	5.92
Within Groups	1,033.59	1,037	1.00	
Total	1,051.28	1,040		
BDS Index				
Between Groups	39.49	3	13.16	22.95
Within Groups	594.76	1,037	.57	
Total	634.25	1,040		
CARS Index				
Between Groups	18.74	3	6.25	9.69
Within Groups	666.59	1,034	.65	
Total	685.33	1,037		

F tests are all significant at the .001 level.

TABLE 3
BDS and CARS Differences for Educational Level and Post Hoc Analyses (Scale Range = 5–9)

<i>Factor</i>	<i>Mean/SD</i>	<i>Bachelor's</i>	<i>Bachelor's Plus</i>	<i>Master's</i>	<i>Postgraduate</i>
Selflessness					
Bachelor's (<i>n</i> = 449)	7.55/1.07				
Bachelor's Plus (<i>n</i> = 339)	7.88/0.69	*			
Master's (<i>n</i> = 155)	7.93/0.49	*			
Postgraduate (<i>n</i> = 98)	7.97/0.62	*			
Integrity					
Bachelor's	7.11/1.15				
Bachelor's Plus	7.54/0.84	*			
Master's	7.79/0.57	*	*		
Postgraduate	8.04/0.54	*	*	*	
Competency					
Bachelor's	7.55/1.03				
Bachelor's Plus	7.85/0.71	*			
Master's	7.92/0.44	*			
Postgraduate	7.98/0.59	*			
Spiritual Appreciation					
Bachelor's	6.38/1.08				
Bachelor's Plus	6.54/0.99				
Master's	6.51/0.86				
Postgraduate	6.83/0.78	*	*	*	
BDS Index					
Bachelor's	7.15/0.93				
Bachelor's Plus	7.45/0.68	*			
Master's	7.54/0.44	*			
Postgraduate	7.71/0.51	*	*	*	
CARS Index					
Bachelor's	6.57/1.01				
Bachelor's Plus	6.75/0.66	*			
Master's	6.81/0.55	*			
Postgraduate	7.00/0.47	*	*	*	

**p* < .05.

and also correlated highly with the CARS Self-Form ratings.

The ANOVA results indicated significant differences between educational level groups for all factors (see Table 2). To establish which groups differed from each other, post hoc analyses (i.e., Newman-Keuls) were performed. These results are provided in Table 3 with the asterisk indicating which educational levels were significantly different (*p* < .05) from each other. The mean values for each educational level for each of the four BDS sub-scales plus the overall BDS index (i.e., average of all 65 items), as well as the CARS Index are also pro-

vided in Table 3. Review of Table 3 indicates that scores on three of the four BDS subscales, the overall BDS Index, and the CARS Index increased with educational level. The one exception was for the Spiritual Appreciation subscale where the master's level subscale score was lower, though not significantly, than the bachelor's plus professional training level subscale score. Examination of the post hoc analyses indicate that the bachelor's educational level differed significantly (*p* < .05) from each of the other educational levels for all but one of the BDS subscales, the overall BDS Index, and for the CARS Index. The one BDS

subscale exception was bachelor's educational level not differing significantly from bachelor-plus and master's degree for the spiritual appreciation BDS subscale. In addition, the bachelor's plus professional training group differed from the postgraduate level for the BDS subscales of integrity and spiritual appreciation as well as for the overall BDS Index and CARS Index. The bachelor's plus professional training group also differed from the master's level for integrity and for the postgraduate level for integrity and spiritual appreciation and for the BDS Index and CARS Index. The master's level and postgraduate level educational groups differed from each other for the BDS subscales of integrity and spiritual appreciation as well as for the overall BDS Index and CARS Index.

Additional ANOVA results associated with the Hypothesis 1, provides further evidence of change from the freshmen to senior years of

the cadet experience (see Table 4). The results of post hoc analyses (i.e., Newman-Keuls) are provided in Table 5 with the asterisk indicating which cadet/graduate level was significantly different ($p < .05$) from the others. Also provide in Table 5 are the mean values of each cadet/graduate group for each of the four BDS sub-scales plus the overall BDS index. Review of Table 5 indicates that scores on two of the BDS sub-scales, and the overall BDS Index increased progressively from the freshmen to senior to graduate levels. One major exception was for the subscale of selflessness that had no significant differences between groups. Another exception was for the subscale of spiritual appreciation that only had significant differences between seniors and freshmen with seniors having higher scores. Considering the data as a whole, seniors and graduates scores were significantly higher than freshmen. In addition, for the subscales of integrity and

TABLE 4
BDS and CARS Differences for USAF Academy Cadets and Graduates

<i>Factor</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Selflessness				
Between Groups	.42	2	.21	00.30 ^{ns}
Within Groups	817.75	1165	.70	
Total	818.16	1167		
Integrity				
Between Groups	61.34	2	30.67	38.42*
Within Groups	929.94	1165	.80	
Total	991.28	1167		
Competency				
Between Groups	10.10	2	5.06	8.19*
Within Groups	718.14	1165	.62	
Total	728.23	1167		
Spiritual Appreciation				
Between Groups	12.61	2	6.31	5.44*
Within Groups	1,349.36	1165	1.16	
Total	1,361.97	1167		
BDS Index				
Between Groups	12.02	2	6.01	12.48*
Within Groups	560.95	1165	.48	
Total	572.96	1167		

ns (not statistically significant)

* $p < .001$.

TABLE 5
BDS Means and Post Hoc Analyses Differences for USAF Academy Cadets and Graduates (Scale Range 5–9)

<i>Factor</i>	<i>Mean/SD</i>	<i>Freshmen</i>	<i>Seniors</i>	<i>Graduates</i>
Selflessness				
Freshmen	7.85/0.96			
Seniors	7.84/0.49			
Graduates	7.89/0.57			
Integrity				
Freshmen	6.87/0.97			
Seniors	7.25/0.89	*		
Graduates	7.41/0.71	*	*	
Competency				
Freshmen	7.63/0.88			
Seniors	7.75/0.76			
Graduates	7.87/0.56	*	*	
Spiritual Appreciation				
Freshmen	6.28/1.11			
Seniors	6.52/1.10	*		
Graduates	6.37/0.97			
BDS Index				
Freshmen	7.15/0.76			
Seniors	7.34/0.69	*		
Graduates	7.39/0.50	*		

* $p < .05$.

TABLE 6
Gender Differences for the BDS and CARS

<i>Factor</i>	<i>Gender</i>	<i>Mean</i>	<i>SD</i>
Selflessness	Male	7.79	.82
	Female	7.97	.68
Integrity	Male	7.18	.93
	Female	7.47	.81
Competency	Male	7.70	.79
	Female	7.85	.67
Spiritual Appreciation	Male	6.31	1.05
	Female	6.73	1.00
BDS Index	Male	7.25	.70
	Female	7.51	.63
CARS Index	Male	6.64	.73
	Female	6.72	.69

All differences significant at the .001 except for the CARS Index which is significant at the .002 level.

competency there were significant differences between the groups of freshmen, seniors, and graduates.

Hypothesis 2, *females will have higher scores than males on the BDS subscales, overall BDS Index, and the CARS Index*, was tested

for differences between the two genders. The basis for including this hypothesis is the research of Rest et al. (1999) who found females scored higher in moral reasoning than males. Results are provided in Table 6. Females scored significantly higher than males

on each BDS subscale ($p < .001$), overall BDS Index ($p < .001$), and the CARS Index ($p < .002$).

The results of the test of Hypothesis 3, *honor representatives will score higher on the BDS subscale of integrity than cadets in general*, are provided in Table 7. There are 164 honor representatives at the academy selected from the senior and junior classes. For this analysis 73 honor representatives volunteered and were administered the BDS. Freshmen and senior participants ($n = 1,051$) were combined to form a group of general cadet scores, representative of the “standard” cadet that was not subjectively identified as an honor representative thought to exude higher qualities of character. Cadets serving as honor representatives scored significantly higher on integrity, the subscale of selflessness and the overall BDS Index. Table 7 indicates that the mean scores for honor representatives were higher than cadets in general.

Hypothesis 4, *the BDS and the CARS will not be subject to participants responding in a socially desirable manner*, was tested by

embedding the Crown-Marlowe Scale (Crown and Marlowe, 1960) within the BDS and CARS. The Crown-Marlowe Scale focuses on ordinary personal and interpersonal behaviors where respondents respond in a socially desirable manner to items on a five-point scale (1 = Strongly Disagree, 3 = Neither, and 5 = Strongly Agree). It was embedded as a potential control for participants answering in a socially desirable way. In turn, the Crown-Marlowe Scale (short form) score was correlated with the overall BDS Index and the CARS Index for a sample of incoming cadets ($n = 143$). The mean value of the Crown-Marlowe Scale was 3.1 on the five-point scale, indicating a midrange score in the “Neither” range. Correlation results indicated the cadets did not answer either the BDS or CARS in a socially desirable manner. The correlations were not significant between the Crown-Marlowe Scale and the BDS Index ($r = -.06$) nor with the CARS Index ($r = .01$). In addition, neither did any of the subscales of the BDS correlate significantly (range of $r = -.02$ to $-.07$) with the Crown-Marlowe Scale. These

TABLE 7
Honor Representatives and Academy Cadets in General (Scale Range 5–9)

<i>Factor</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>
Selflessness			
Honor Representatives	7.95	.46	3.73***
Cadets in General	7.72	.88	
Integrity			
Honor Representatives	7.18	.73	2.29*
Cadets in General	6.97	.92	
Competency			
Honor Representatives	7.64	.60	.57
Cadets in General	7.59	.81	
Spiritual Appreciation			
Honor Representatives	6.44	1.04	1.31
Cadets in General	6.28	1.06	
BDS Index			
Honor Representatives	7.31	.49	2.63**
Cadets in General	7.14	.72	

df = 1,122

* $p < .02$; ** $p < .01$; *** $p < .001$.

results are consistent with those of Barlow (2002) who also correlated the BDS and CARS indexes with the Crown-Marlowe Scale for a separate sample of Air Force officers ($n = 346$) in educational programs at the Air University. She reported the mean value of the Crown-Marlowe Scale to be 2.7 with a coefficient alpha of .92 for the scale. She also found that the Crown-Marlowe Scale did not significantly correlate with the BDS or CARS scales.

Convergent and Divergent Validation

Hypothesis 5 and 6 are associated with establishing convergent and divergent validity of the two assessment instruments (i.e., BDS and CARS). Hypothesis six, *the BDS Index will be correlated positively with the CARS*, was tested using correlational analysis. Logically, one would expect convergence if similar concepts associated with character were significantly correlated and those that are dissimilar to character are not correlated. Results of this analysis gave partial support for Hypothesis 6. That is, the BDS Index correlated significantly with the CARS Index ($r = .43$, $n = 4,158$, $p < .001$).

The Hypothesis 6, *the BDS and CARS Indices will be positively correlated with the moral reasoning indices of the DIT*, was also tested using correlational analysis. There is little reason to think that there would be a strong relationship between dimensions of moral excellence and the DIT's measures of moral

reasoning (P% and N2 Indices). Nevertheless, one would hope that there would be a small positive relationship between the BDS and CARS indices and the DIT's P% Index and N2 Index. Analysis results are provided in Table 8. The data show a pattern of low but significant correlations between the BDS Index and the P% Index and N2 Index with the exception for the BDS subscale of competency which was not significantly correlated with the P% Index. The CARS Index also had a low but significant correlation with the N2 Index but not with the P% Index. Therefore, these results show a low correlation between the BDS and CARS Indices with the DIT-2 measures of moral reasoning.

Summary

This study, Study 2, described the process used and results obtained while validating the two character assessment instruments whose development was discussed in Study 1. Although this study provides results suggesting that these instruments are valid in terms of discriminant, convergent, and divergent validity, these results need to be replicated with a different sample in a different setting. Study 3 provides additional support by replicating some of the results of this study. Specifically, Study 3 focuses on performing confirmatory factor analysis for both the BDS and CARS using different subjects in a different setting. Furthermore, Study 3 examines whether or not

TABLE 8
BDS, CARS, and DIT-2 Correlations

<i>Scales and Subscales</i>	<i>DIT P% Index</i>	<i>DIT N2 Index</i>
BDS ($n = 374$)	.16**	.15**
Selflessness	.16**	.16**
Integrity	.12*	.16**
Competency	.09	.10*
Spiritual Appreciation	.13*	.08
CARS ($n = 358$)	.01	.12*

* $p < .05$; ** $p < .01$.

those completing the BDS and CARS acted in a socially desirable manner.

STUDY 3: RESEARCH REPLICATION

Method

Participants

Participants consisted of 1,114 active duty Air Force Officers attending education and training programs at the Air Force's Air University. The number of cases for each educational program and average age was: Aerospace Basic Course, 514 participants, average age 26; Squadron Officers School, 346 participants, average age 30; Air Command and Staff College, 106 participants, average age 37; Air War College, 50 participants, average age 43; and Senior Leader Orientation Course, 98 participants, age not collected. The overall age ranged from 20 to 54 years with an average age of 29.6 years. The racial composition was: American Indian or Alaskan Native, 1.1%; Asian or Pacific Islander, 3.6%; Black, not of Hispanic Origin, 8.0%; Hispanic, 4.9%; White, Not Hispanic Origin, 77.7%; Other, 4.7%. For the total sample, 72.6% were males and 27.4% were females.

Measures and Procedure

The BDS and the CARS were the primary instruments used in the confirmatory factor analysis portion of this research. In addition, the Crown-Marlowe (Crown & Marlowe, 1960) short form social desirability scale was used to check for participants responding in a socially desirability fashion.

Analysis and Results

Barlow, Jordan, and Hendrix (2003), performed research to replicate the survey factor

structure reported above in Study 2. Their replication involved performing confirmatory factor analysis (Arbuckle & Wothke, 1999) for both the BDS as well as the CARS. For the BDS, Barlow et al. (2003) used the 65-item version of the BDS and conducted confirmatory factor analysis by imposing the factor model described in research Study 2 above and then established the goodness-of-fit of their data to the model. The resulting analysis confirmed that the data had a good fit to the model giving additional support for the factor structure reported in study 2 above. Specifically, the goodness-of-fit indexes for the BDS subscales were: normed fit index = .94, relative fit index = .93, incremental fit index = .94, Tucker-Lewis index = .94, comparative fit index = .94. The coefficient alpha reliabilities for the four factors of the BDS were within .02 units (.88 to .94) of those reported in Study 2 above. These were: Selflessness (coefficient alpha = .94); Integrity (coefficient alpha = .89); Competency (coefficient alpha = .88); and Spiritual Appreciation (coefficient alpha = .88). Barlow et al. (2003) also performed confirmatory factor analysis with data collected using the CARS and found that the data also had a good fit to the single factor reported in Study 2 above for the CARS. The goodness of fit indexes for the CARS instrument were normed fit index = .98, relative fit index = .98, incremental fit index = .99, Tucker-Lewis index = .98, comparative fit index = .99. Barlow et al. (2003) added the Crown and Marlowe (1960) short-form social desirability scale to the BDS and CARS for one of their participant groups ($n = 346$). They reported that there was no indication that participants answered in a socially desirable manner. The Crown-Marlowe Scale did not correlate significantly with the BDS subscale scores (Selflessness = $-.10$, ns; Integrity = $-.04$, ns; Competency = $-.12$, ns; and Spiritual Appreciation = $-.11$, ns) nor did it correlate with the CARS ($-.05$, ns). In addition, neither did the Crown-Marlow Scale make a significant contribution to the explained variance in the BDS model when

entered as a control variable using hierarchical regression ($F = .20, p = .85$).

DISCUSSION AND CONCLUSIONS

This research has practical and theoretical implications. A great deal of research has been performed dealing with moral reasoning, but little has been carried out with the broader construct of moral excellence or moral character (Walker & Pitts, 1998). In addition, there does not seem to be a set of well-developed measurement instruments that focus on the concepts of moral excellence. This research addressed this practical issue of assessing character assessment or moral excellence. It involved developing two character assessment instruments founded on various dimensions of moral excellence. From a theoretical perspective, this research builds on, and is consistent with, the theoretical models of the moral person proposed by Lickona (1991) and Berkowitz (2002) that have cognitive, affective, and behavioral components in common.

The studies cited in this report dealt with the development and validation of two instruments designed to measure the construct of moral character or moral excellence. The hypotheses focused on discriminant validity and, to a lesser extent, convergent and divergent validity. Discriminant validity was established with a variety of different groups in different settings. These included cadets at the USAF Academy, Air Force Academy graduates 3 years after graduation, and Air Force Officers attending professional education programs. Results consistently indicated that the BDS and the CARS Self-Form distinguished between these groups. The significant differences were consistent with DIT findings (Rest & Narvaez, 1994; Rest et al., 1999) and character-related literature (e.g., Astin, 1993; Myrsky & Helkama, 2002; Pascarella & Terenzini, 1991).

Together these three studies provide a basis for evaluating the utility of the instruments in academic settings. The Behavioral Desirability

Scale was designed to measure one aspect of the cognitive domain, that of individual character-related values. This 65-item instrument contained four factors: integrity, selflessness, competency, and spiritual appreciation. The CARS was designed to measure one aspect of the behavioral domain, that of character-related behaviors of individuals. A Behavioral Observation Rating Scale (BORS) format was used for rating 12 dimensions of character: integrity, honesty, selflessness, compassion, competency, respectfulness, fairness, responsibility and self-discipline, decisiveness, spiritual appreciation, and cooperativeness.

The results of this research suggest that the BDS is a useful addition for measuring character-related values associated with moral excellence. Furthermore, the CARS Self-Form, even though it is a self assessment instrument, appears to also be a useful instrument when combined with the BDS for measuring changes in character. Even though the sample size of the 360-degree CARS assessment was small, it was encouraging to note that the correlations between supervisor, peer, and subordinate ratings correlated highly with those of the self ratings (ranging from .83 to .90). The CARS also seems to be a useful instrument for rating group members' supervisors or leaders. The CARS appears to be a useful device for academic and business settings.

Strengths of this research include the sample's composition and size for the BDS and CARS Self-Form, the internal consistency reliability of the four BDS scales, the absence of social desirability effects on the BDS and CARS Self-Form, the stable factor structure for the BDS and CARS, and the capacity of both instruments to differentiate between groups. This research yielded increasing scores from new cadets to senior level officers. The BDS and CARS scores increased with age and educational level. The BDS was found to correlate significantly with rated character using the CARS Self-Form and when rated by others. The results indicated that scoring high on leadership and character had a positive

effect on organizational outcomes. This result supports Emler and Cook's (2001) and Kouzes and Posner (1987) emphasis on character or integrity as one of the most important factors for leadership effectiveness.

One limitation of this research is the lack of adequate data for analyzing the 360-degree CARS process. Future research is planned to increase the sample size in order to establish the utility of this instrument. Another limitation involves the composition of the samples in this research. Even though it involved a large sample with many occupations included, the sample was mainly Air Force personnel. Future research should broaden the sample composition and include other civilian professions such as business, law, medicine, and education.

Further research is necessary at institutions with specific interventions intended to develop character within its students to clearly link the impact of deliberate character education to the BDS and CARS. Additionally, research on the relationship between the BDS and CARS with education is necessary.

One conclusion is that the BDS and CARS Self-Form seem to be measuring a broader or different aspect of moral development than the DIT, an instrument that is well suited for measuring moral reasoning. The character assessment instruments described in this research might provide additional utility, when combined with the DIT, in measuring dimensions of moral development not measured by the DIT.

APPENDIX

Behavioral Desirability Scale Factors, Factors Loadings, and Coefficient Alpha Reliabilities
Factor 1: Selflessness (Respectfulness, Compassion, Loyalty, and Cooperativeness) (Alpha = .93)

<i>Item</i>	<i>Factor Loading</i>
1. A person is known for being there for his friends.	0.69
2. An individual can be depended upon to help his neighbors.	0.73
3. A boss lets his workers know that he cares about them.	0.64
4. A person, by her actions, shows that she respects every individual and feels all have intrinsic worth.	0.63
5. People of a different nation and different continent are suffering. A man gives of his time to raise money so that the people might have a better life.	0.65
6. After working an eight-hour day, a couple consistently finds time to help their three children with homework.	0.67
7. A worker acts in a manner that shows he cares about the life of the company.	0.69
8. No matter how busy, he is a worker that always finds time to help with group projects.	0.56
9. Even though a man has to stay late, he makes sure the equipment is ready for a colleague's presentation the next morning.	0.72
10. A couple baked cookies for new next-door neighbors to welcome them to their neighborhood.	0.66
11. A person, when faced with a series of choices, makes a choice that is best for the majority of his associates even though it costs him personally.	0.53
12. A man takes time out of his busy schedule to listen to a friend's problems.	0.67
13. Even though a man's job is very important to his family's livelihood, he makes time for his duties at home as a father and husband.	0.65
14. Even though it makes a young man feel extremely uncomfortable to be around a friend's dying mother, he still goes with his friend to the hospital.	0.60
15. A person gives up his seat on a bus to a man who is older and has a physical handicap.	0.59
16. A head manager is very busy but she always makes time to speak with employees that are having problems.	0.61
17. When faced with a situation that may give personal benefit, a worker puts the interest of his organization and the welfare of others first.	0.51

Behavioral Desirability Scale Factors, Factors Loadings, and Coefficient Alpha Reliabilities
Factor 2: Integrity (Honesty) (Alpha=.87)

<i>Item</i>	<i>Factor Loading</i>
1. A man is not sick but takes two days of sick leave since he will lose it if he doesn't take the leave by month's end.	0.55
2. A teenager tells her little sister to ignore their mother's punishment of not leaving the house and go out. The teenager says, "I'll cover for you."	0.47
3. A student, taking a very hard class in which many students are cheating, gets and uses test answers that others have obtained.	0.45
4. When a man sees a line in front of his favorite restaurant he tries to find a way to get to the front without waiting.	0.47
5. A person uses other people's ideas and influence to get ahead.	0.41
6. An individual calls in sick because there is a sale at the mall.	0.46
7. A woman is more concerned with pleasing her boss than making an informed decision about the personnel changes needed in the office.	0.42
8. A man feels that his receiving a bigger pay raise than his coworkers, whose performance is the same as his, is justified since he makes a point of getting along with their boss.	0.41
9. A coworker tells his friends that his boss is a jerk. This happens after the boss corrects him for a mistake he made.	0.41
10. When presented with a complex problem a worker gives it to another worker to solve.	0.37
11. In order to get out of jury duty, a person claimed he knew the defendant even though he had never seen him before in his life.	0.38
12. A worker accepts credit for work his coworkers helped him to complete.	0.31
13. When filling out tax forms a person overstates the amount she has given to charities.	0.39
14. An individual calls his company's finance office to inform them of an extra \$100 he received during his last pay period.	-0.48
15. A worker turned in her coworker for a serious company infraction, even though they enjoyed a close relationship.	-0.39
16. A man tends to tell the truth only when there is no negative consequence to him.	0.32
17. A woman tells a cashier that he gave her too much change.	-0.28
18. A person is consistent in doing the right thing even when faced with less ethical choices which may profit him or her.	-0.30

Behavioral Desirability Scale Factors, Factors Loadings, and Coefficient Alpha Reliabilities
Factor 3: Competency (Responsibility and Fairness) (Alpha=.88)

<i>Item</i>	<i>Factor Loading</i>
1. A worker consistently produces top quality work.	-0.36
2. A father, who is asked for help with homework by his child, makes sure he gets home from work on time so he can help the child.	-0.34
3. A supervisor listens to everyone's ideas and sets up a clear training plan for the entire company.	-0.44
4. A person promises his friends he will get a designated driver for the party but never does.	0.45
5. A woman tends to put off work-related tasks until the last minute.	0.38
6. Individuals trust their boss since he is open in his conversation with all individuals.	-0.45
7. A young man is somewhat passive; he lets others run his life.	0.50
8. A man tends to make up excuses to himself for his failures.	0.50
9. An employee is usually late for meetings.	0.54
10. When given any task, the person excels.	-0.32
11. Even though it is difficult, a boss tells her employees when they have not performed well.	-0.34
12. A person lives within his financial means.	-0.32
13. A young man is exceptionally capable and uses his skills for the benefit of his organization and coworkers.	-0.68

14. A manager, when conflict arises among his employees, makes an effort to reach a peaceful resolution. -0.62
15. A person is extremely competent and excels in all assignments. -0.65

Behavioral Desirability Scale Factors, Factors Loadings, and Coefficient Alpha Reliabilities
Factor 4: Spiritual Appreciation (Alpha = .87)

Item	Factor Loading
1. A person is an agnostic, but appreciates others spiritual beliefs and doesn't impose his beliefs and views on others.	0.65
2. A person, after studying a wide range of religions and their doctrines, has decided that she is an agnostic.	0.60
3. A person attends religious ceremonies other than his faith when given the opportunity.	0.67
4. A person of faith is willing to share his views if asked but doesn't impose his views on others.	0.48
5. A woman tells a man that he is going to "Hell" because he doesn't believe in her religious values and beliefs.	-0.64
6. A woman considers herself as very religious and prefers to learn more about her faith and not about the faith of others.	-0.47
7. A woman grew up as a Christian and feels it is her place to convert those of other faiths to Christianity.	-0.65
8. A woman feels strongly about her spiritual views and tries to get others to subscribe to her views and beliefs.	-0.58
9. A father is upset with his son for showing an interest in Buddhism while the rest of the family is devout Christians.	-0.69
10. An organization supports individuals' rights to differ in their spiritual beliefs.	0.41
11. A group of individuals, with different spiritual values, get together to share their spiritual views and beliefs.	0.41
12. A person enjoys learning about other religions, including their differences and similarities among them.	0.56
13. An organization appreciates individuals of different spiritual backgrounds and interests.	0.47
14. A person makes a point of reading about other faiths simply to understand the different views others have.	0.39
15. An individual has a variety of friends with different religious backgrounds.	0.44

Author Note: Behavioral Desirability Scale, Copyright © 2001, William H. Hendrix.

REFERENCES

- Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action-control: From cognition to behavior* (pp. 11-39). Heidelberg, Germany: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Astin, A. W. (1993). *What matters in college?* San Francisco: Jossey-Bass.
- Battistich, V. (1999, November). *Toward a more adequate assessment of moral development*. Paper presented at the meeting of the Association for Moral Education, Minneapolis, MN.
- Battistich, V., Solomon, D., Watson, & Schaps, E. (1997). Caring school communities. *Educational Psychology*, 32, 137-151.
- Barlow, C. B. (2002). *Character assessment across Air Force professional military education: A descriptive investigation*. (Res. Rep. No. AU/

- ACSC/015/2002-4). Air University, Maxwell AFB, Alabama.
- Barlow, C. B., Jordan, M., & Hendrix, W. H. (2003). Character assessment: An examination of leadership levels. *Journal of Business and Psychology, 17*(4), 563-584.
- Bebeau, M. J. (2002). The Defining Issues Test and the Four Component Model: Contributions to professional education. *Journal of Moral Education, 31*(3), 271-295.
- Berkowitz, M. W. (1995). *The education of the complete moral person*. Aberdeen, Scotland: Gordon Cook Foundation.
- Berkowitz, M. W. (2002). The science of character. In W. Damon (Ed.), *Bringing in a new era in character education* (pp. 43-63). Stanford, CA: Hoover Institute Press.
- Colby, A., Kohlberg, L., Gibbs, J., & Lieberman, M. (1983). A longitudinal study of moral judgment. *Society for Research on Child Development Monograph, 48*(1-2, Serial No. 200).
- Crowne, D. P., & Marlowe, D. A. (1960). New scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.
- Davidson, M. L., & Khmelkov, V. T. (2003). *Global portraits of social and moral health—For youth and adults*. Cortland NY: Center for the 4th and 5th Rs.
- Eiferman, D. (1982). *Moral judgment in adult urban college students*. Paper presented at the East Coast Forum on Urban Higher Education Research, New York.
- Eisenberg, N. (1995). Prosocial development: A multifaceted model. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Moral development: An introduction* (pp. 401-429). Boston: Allyn & Bacon.
- Emler, E., & Cook, T. (2001). Moral integrity in leadership: Why it matters and why it may be difficult to achieve. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 277-298). Washington, DC: American Psychological Association.
- Flannery, B. L., & May, D. R. (2000). Environmental ethical decision making in the U.S. metal-finishing industry. *Academy of Management Journal, 43*, 642-662.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Kleiner, C., & Lord, M. (1999, November 22). The cheating game. *U.S. News & World Report, 127*, 55-66.
- Kohlberg, L. (1981). *Essays on moral development: Vol. 1. The philosophy of moral development*. San Francisco: Harper & Row.
- Kohlberg, L. (1984). *The psychology of moral development (Vol. 1)*. San Francisco: Harper & Row.
- Kohlberg, L., & Higgins, A. (1987). School democracy and social interaction. In W. Kurtines & J. Gewirtz (Eds.), *Moral development through social interaction* (pp. 102-128). New York: Wiley & Sons.
- Kouzes, J. M., & Posner, B. Z. (1987). *The leadership challenge: How to get extraordinary things done in organizations*. San Francisco: Jossey-Bass.
- Lei, T. (1981). *The development of moral, political, and legal reasoning in Chinese societies*. Unpublished master's thesis, University of Minnesota, Minneapolis and St. Paul.
- Lei, T., & Cheng, S. (1984). *An empirical study of Kohlberg's theory and scoring system of moral judgment in Chinese society*. Unpublished manuscript, Harvard University, Center for Moral Education, Cambridge, MA.
- Lemming, J. S. (1993). In search of effective character education. *Educational Leadership, 51*, 63-71.
- Likona, T. (1991). *Educating for character*. New York: Bantam.
- Lind, G. (1985). Moral competence and education in democratic society. In G. Zecha & P. Weingartner (Eds.), *Conscience: An interdisciplinary view*. Dordrecht, The Netherlands: Reidel.
- Lind, G. (1986). Growth and regression in cognitive-moral development. In C. Harding (Ed.), *Moral dilemmas: Philosophical and psychological issues in the development of moral reasoning*. Chicago: Precedent.
- Loxley, J., & Whiteley, H. (1986). *Character development in college students (Vol. 2)*. Schenectady, NY: Character Research Press.
- Mentkowski, M., & Strait, M. (1983). *A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum* (Final report to the National Institute of Education, Research Report No. 6). Milwaukee, WI: Alverno College, Office of Research and Evaluation.
- Myrsky, L., & Helkama, K. (2002). The role of value priorities and professional ethics training in moral sensitivity. *Journal of Moral Education, 31*(1), 35-50.

- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass.
- Rest, J. R. (1983). Morality. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol 3. Cognitive development* (pp. 556-629). New York: Wiley.
- Rest, J. R., & Narvaez, D. (Eds.). (1994). *Moral development in the professions: Psychology and applied ethics*. Hillsdale, NJ: Erlbaum.
- Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Erlbaum.
- Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). Chicago: University of Chicago Press.
- Vessels, G. G. (1998). *Character and community development: A school planning and teacher training handbook*. Westport CT: Praeger.
- Walker, L. J., & Pitts, R. C. (1998). Naturalistic conceptions of moral maturity. *Developmental Psychology*, 34, 403-419.
- Whitla, D. (1978). *Value added: Measuring the impact of undergraduate education*. Cambridge, MA: Harvard University, Office of Instructional Research and Evaluation.