

THE PSYCHOMETRIC ASSESSMENT OF MEANING-MAKING:
REACTIONS TO EVERYDAY DILEMMAS (RED)

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

Robin S. Salter

Certificate of Approval:

Virginia E. O'Leary
Professor (Retired)
Psychology

Philip M. Lewis, Chair
Professor
Psychology

William F. Buskist
Professor
Psychology

Charlotte D. Sutton
Associate Professor
Management

Joe F. Pittman
Interim Dean
Graduate School

THE PSYCHOMETRIC ASSESSMENT OF MEANING-MAKING:
REACTIONS TO EVERYDAY DILEMMAS (R.E.D.)

Robin S. Salter

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
December 15, 2006

THE PSYCHOMETRIC ASSESSMENT OF MEANING-MAKING:
REACTIONS TO EVERYDAY DILEMMAS (RED)

Robin S. Salter

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon request of individuals or institutions and at their expense. The author reserves all publication rights.

Signature of Author

Date of Graduation

VITA

Robin Seely Salter, daughter of Bob and Nancy Seely, was born on May 23, 1961, in Rock Hill, South Carolina. She graduated from Georgia Southern College with a Bachelor of Business Administration in Marketing in June, 1984. After spending most of her early career in marketing research, she returned to Georgia Southern and completed a Masters of Science degree in Psychology in August, 1998. She entered the Auburn University Industrial/Organizational Psychology program in August, 1998. During her enrollment at Auburn University her applied training included a two-year fellowship with the Army Research Institute at Fort Benning, Georgia through the Consortium Research Fellows Program, and a one-year internship with the Alabama State Personnel Department. She married Armin Salter on August 10, 1985 and they have two children, Rachel, 19 and Robby, 17.

DISSERTATION ABSTRACT
THE PSYCHOMETRIC ASSESSMENT OF MEANING-MAKING:
REACTIONS TO EVERYDAY DILEMMAS (RED)

Robin S. Salter

Doctor of Philosophy, December 15, 2006
(M.S., Georgia Southern University, 1998)
(B.B.A., Georgia Southern College, 1984)

106 Typed Pages

Directed by Philip M. Lewis

According to Robert Kegan (1994), many people are essentially “in over their heads” in trying to meet the conceptual and emotional demands of everyday life. In the workplace, this phenomenon may be exacerbated if jobs are assigned that exceed an individual’s meaning-making capacity, mentors are poorly matched with protégés, or employees are sent through training programs before they are developmentally ready to assimilate the training experience. Therefore, a convenient measure assessing the developmental level of employees could be of benefit to organizations. The purpose of the present study was to test and validate a new measure, *Reactions to Everyday Dilemmas* (RED), which estimates an individual’s stage of “meaning-making” as defined by Kegan’s (1982, 1994) constructive-developmental theory of personality. According to Kegan, individuals progress through a series of hierarchically ordered stages during

which their frameworks for interpreting experiences become increasingly complex. As people advance from one stage to the next, a more encompassing meaning-making framework influences how they react to complex situations at work, school, and in their personal lives. The present paper describes two studies to establish RED's reliability and construct validity. RED was compared with scores from the *Defining Issues Test* (DIT) (Rest, 1975, 1979) to establish convergent validity, and the *Life Orientation Test* (LOT-R) (Scheier, Weintraub, & Carver, 1986; Scheier, Carver, & Bridges, 1994) to assess discriminant validity. Modest significant correlations were observed between several of RED's stage scales and those of the DIT, even after controlling for the covariates of age and education. Two of RED's scales also correlated with the LOT-R. The LOT-R unexpectedly correlated with age and educational level, and the relationship between RED and the LOT-R decreased substantially after controlling for these two variables. Suggestions for future revisions of the RED and research steps are discussed.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Phil Lewis for his support throughout the lengthy process of developing, revising, and reworking a new assessment. His expert contributions were extremely valuable to the development of this assessment. Many thanks to Drs. Bill Buskist, Virginia O’Leary, and Charlotte Sutton for the guidance they provided as members of the dissertation committee. I would also like to thank Dr. Randolph Pipes for serving as the outside reader.

I would like to thank my husband, Armin Salter, whose unwavering confidence and support made this dissertation possible. Thank you for agreeing to move from Savannah to Auburn, and for understanding when it became evident that we would probably be here for more than a couple of years. Many thanks to my family, Armin, Rachel and Robby Salter, and Bob, Nancy, and Pam Seely, for encouraging me to keep pushing forward. Lastly, I would like to express my appreciation to Ed, Achilles, and Mimo for providing good company when I was burning the midnight oil.

This dissertation was written according to the 2001 Publication Manual of the American Psychological Association, Fifth Edition.

Microsoft Word 2003 was used for word processing, and data analyses were performed with SPSS 14.0.

TABLE OF CONTENTS

LIST OF TABLES.....	x
I. INTRODUCTION.....	1
II. STUDY 1.....	30
Method	30
Results	32
Study 1 Discussion.....	41
III. STUDY 2.....	44
Method	44
Results	48
Study 2 Discussion	71
IV. GENERAL DISCUSSION.....	72
REFERENCES.....	78
APPENDICES.....	86
Appendix A: Item Rating Form for the Reactions to Everyday Dilemmas Meaning-Making Assessment	87
Appendix B: Reactions to Everyday Dilemmas Assessment for Study 1	89
Appendix C: Reactions to Everyday Dilemmas Assessment for Study 2....	92

LIST OF TABLES

TABLE 1.	Summary of Study 1 Sample Characteristics.....	30
TABLE 2.	Study 1 Mean Ranking for Each Item by Story in the RED Measure	34
TABLE 3.	Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 2 Scale	35
TABLE 4.	Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 3 Scale	36
TABLE 5.	Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 4 Scale	37
TABLE 6.	Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 5 Scale	38
TABLE 7.	Study 1 Alpha Coefficients after Removing Cases and Items for the RED Measure	39
TABLE 8.	Study 1 Bivariate Correlations among the RED Stage Scales, Age, Education and College Class.....	40
TABLE 9.	Summary of Study 2 Sample Characteristics.....	44
TABLE 10.	Percentage of Participants Who Failed to Complete Specific Portions of the RED Measure in Study 2	50
TABLE 11.	Study 2 Mean and Median Stage Level Scores for Each Scenario from the RED Measure	51
TABLE 12.	Study 2 Stage Level Rating Scores for Story A from the RED Measure (Frequencies and Percentages)	52

TABLE 13.	Study 2 Stage Level Rating Scores for Story B from the RED Measure (Frequencies and Percentages)	53
TABLE 14.	Study 2 Stage Level Rating Scores for Story C from the RED Measure (Frequencies and Percentages)	54
TABLE 15.	Study 2 Stage Level Rating Scores for Story D from the RED Measure (Frequencies and Percentages)	55
TABLE 16.	Study 2 Stage Level Rating Scores for Story E from the RED Measure (Frequencies and Percentages)	56
TABLE 17.	Study 2 Stage Level Ranking Scores for Story A from the RED Measure (Frequencies and Percentages)	57
TABLE 18.	Study 2 Stage Level Ranking Scores for Story B from the RED Measure (Frequencies and Percentages)	58
TABLE 19.	Study 2 Stage Level Ranking Scores for Story C from the RED Measure (Frequencies and Percentages)	59
TABLE 20.	Study 2 Stage Level Ranking Scores for Story D from the RED Measure (Frequencies and Percentages)	60
TABLE 21.	Study 2 Stage Level Ranking Scores for Story E from the RED Measure (Frequencies and Percentages)	61
TABLE 22.	Study 2 Comparisons of Cronbach Alpha Reliabilities of Stage Scales Before and After Removing Cases and Items for the RED Measure	62
TABLE 23.	Comparison of Study 2 Defining Issues Test (DIT) Scores with Normative Data	65
TABLE 24.	Study 2 Bivariate Correlations among Rating and Ranking Scales for the RED Measure, DIT, Age, Education, College Class, and LOT-R	68
TABLE 25.	Correlation Coefficients between the RED Ranking Scales, DIT, and LOT Before and After Controlling for Age and Education	70

I. INTRODUCTION

The world of work today requires an ability to handle more than just task-related assignments. As team members, supervisors and subordinates, employees deal with a variety of interpersonal situations that can increase in complexity as they move up the hierarchy of supervisory and management positions. These situations can include being a project leader on a team comprised of peers, supervising employees who were formerly co-workers and friends, advancing from a technical job to a management position, and trying to make sense of organizational change and politics. In any single work environment, employees and supervisors may have to handle conflict, negotiate, persuade, follow orders, give orders, take disciplinary action, appraise performance, receive and respond to their own appraisals, and compete for limited resources in the form of promotions, bonuses, and pay increases.

Cognitively and emotionally, are most of us equipped to meet these challenges? According to Robert Kegan, the answer is no. Kegan (1994) proposed that today's complex world (whether at work or at home) often demands an ability to make meaning of our experiences that exceeds the capacity of most individuals. As a result, people are essentially "in over their heads" in trying to meet the demands of everyday life. The purpose of this research was to develop and test an efficient, quantitative instrument that could identify the complexity of a person's meaning-making framework. As a development tool, this instrument could provide employees and managers with a better

understanding of the lens through which they and others view the world. This improved understanding could help managers and employees clarify role expectations, understand motivations, identify strengths and limitations, and facilitate development without thrusting employees into an “in over our heads” work context.

The new measure, *Reactions to Everyday Dilemmas* (RED), was designed to estimate an individual’s stage of “meaning-making” as defined by Robert Kegan’s (1982, 1994) constructive-developmental theory of personality. According to Kegan, individuals advance through as many as six stages throughout the lifespan during which their framework for interpreting experiences changes qualitatively. As people move from one stage to the next, an increasingly complex meaning-making framework determines how one constructs an understanding of circumstances at work, at school, and in one’s personal life. Currently the only way to assess an individual’s meaning-making stage is by use of a complicated, time-consuming, face-to-face interview (Lahey, Souvaine, Kegan, Goodman, & Felix, 1988). Therefore, organizations would benefit significantly from a paper and pencil assessment that could more efficiently identify an employee’s meaning-making framework. With knowledge of how an employee constructs meaning, employers would be in a better position to mentor, make job assignments and design effective employee development plans tailored to each employee’s developmental level (Forsythe, Snook, Lewis, & Bartone, 2002). Furthermore, identifying an individual’s Kegan stage could help leaders predict how subordinates will respond to certain styles of leadership, as well as predicting the leadership styles that individuals are likely to adopt for themselves (Kuhnert & Lewis, 1987).

In recent years, Kegan and other researchers have begun to emphasize the importance of the level of sophistication with which leaders and managers interpret experience. For example, Kegan and Lahey (2001a, 2001b) have developed an approach for individuals to challenge their “big assumptions” and identify competing commitments. Other leadership program developers are now calling attention to the need to enhance *meaning-making* or *sense-making* among people in work groups to improve collaboration, responses to complex challenges, and adaptability to organizational change (Palus, Horth, Selvin, & Pulley, 2003; Leonard, 2003; Senge, 2005). Palus et al. (2003) have suggested that the demands placed on today’s leaders may require them to operate from Kegan’s highest meaning-making framework, Stage 5. This suggestion, if true, would indicate that the discrepancy between where people are developmentally, and where they need to be has surpassed the “in over our heads” phenomenon that Kegan first proposed in 1994. While training methods are being created to include activities that will enhance meaning-making development, the extent to which participants are assessed in advance to identify their baseline stage of development is not clear. In a recent meta-analysis assessing the effectiveness of managerial leadership development programs, Collins and Holton (2004, p. 240) stated that, “leadership development programs will produce substantial results, especially if they offer the right development programs for the right people at the right time.”

At present, a person’s Kegan stage is typically assessed by conducting a 90-minute interview, and a substantial amount of training is required to conduct and score these interviews. The subject-object interview (Lahey et al., 1988) is a procedure in which individuals are asked to discuss recent events that affected them emotionally.

During the interview, the assessor attempts to identify how interviewees organize these emotional experiences in order to determine what is “subject” versus what is “object” for them. In the subject-object analysis, subject is the organizing framework and object is what is being organized. Subject-object interviewing requires extensive training and practice, and interviews must be transcribed and then scored by trained assessors. Scoring a single interview typically takes several hours. As a result, attempting to incorporate subject-object interviews into a leadership development program could be considered impractical.

Employees might also be uncomfortable with the very personal and intimate nature of the subject-object interview as a career development assessment tool. Because the interviewees are asked to describe recent emotional personal events, subject-object interviews could have a clinical feel and be considered an invasion of privacy if they were incorporated into an employee development process. Tests that feel like clinical assessments are typically considered intrusive and offensive within an employment setting (Jones, 1991). However, if personality measures are paired with ability tests (Rosse, Miller, & Stecher, 1994) or if the personality scales are perceived as being job related (Shuler, 1993), then they are more likely to be considered acceptable within a work context.

One program that integrates developmental assessment with executive coaching is Laske’s (2000) developmental coaching method that begins with an evaluation using the Developmental Structure/Process Tool (DSPT). The DSPT is a two-part interview procedure that includes the subject-object interview (Lahey et al., 1988) to identify the executive’s meaning-making framework and a “dialectical-schema interview”

(Basseches, 1984) to obtain a mental process profile (how an individual processes information while viewing the world through a particular Kegan stage). Laske's developmental coaching program also takes into consideration the meaning-making framework of the executive coach, proposing that coaches can only be effective if their developmental level is above that of the client (Laske, 2004).

Programs such as the DSPT may be effective and appropriate for situations in which organizations are committed to providing the extensive resources required for personalized executive coaching of high-level organizational members. However, managers and employees at levels throughout an organization may be candidates for development programs or special assignments, and what is needed is a convenient, non-intrusive assessment with face validity to identify a person's level of meaning-making. The experimental measure tested in the present research asked respondents to indicate how they would respond to each of five hypothetical dilemmas. Of the five dilemmas, four described situations relevant to the workplace, and the fifth dilemma dealt with a personal relationship problem. Participants were scored by the structural complexity of issues they identified as most significant or relevant to them in each dilemma. If validated successfully, RED could contribute to the advancement of academic research on constructive-developmental theory. This measure might also serve as an employee development assessment tool that would provide organizations with information that could improve decisions about job assignments, enhance team building, and improve mentorship assignments.

Overview of Kegan's Constructive-Developmental Theory

Kegan's (1982, 1994) Constructive/Developmental Theory is based on the premise that (a) individuals construct reality (i.e. make meaning of their experiences) and (b) the framework individuals utilize for meaning-making evolves throughout the lifespan. Kegan proposed that individuals use a conceptual system to derive meaning from internal and external experiences. This system is also described as a set of progressively more complex organizing principles that change qualitatively from one stage to the next (Lahey et al., 1988). Kegan identified six stages, four of which occur during adolescence and adulthood. Individuals spend much of their lives in transition between stages, and constructive-developmental researchers believe that the move from one stage to the next is driven by experiences too complex to be assimilated to the individual's current level of meaning-making.

Each meaning-making stage is defined largely by what individuals are "subject to" (i.e., embedded in) versus what is object for them (i.e., on what they can reflect) (Lahey et al., 1988). As individuals move from one stage to the next, what was subject in the previous stage becomes object in the new stage. RED was designed to assess the meaning-making level of adults, therefore only the four adult stages (2, 3, 4, and 5) will be assessed using the RED instrument. However, descriptions of all six stages are provided below, beginning with the childhood stages, Stage 0 and Stage 1.

Stages 0 and 1: The Incorporative and Impulsive Stages. During the first 18 months or so of infancy, Kegan proposed that the infant operates from a stage that he called *incorporative*, or Stage 0. During this period, children are described as being subject to reflexes (sensing, movement) and are not capable of recognizing other persons

or things as objects distinct from themselves. As a result, infants in this stage are not thought to experience separation anxiety when the caregiver is not present, and quickly lose interest in a toy or a ball that is suddenly removed or covered because they do not view other objects as things to keep or to miss (Kegan, 1982). When the child moves to Stage 1, the *impulsive* stage, reflexes, movements, and physical entities become object (that is, the child can reflect upon them instead of being “embedded” in them). At this point, the child’s framework for organizing the world of physical objects and movement consists of her or his impulses, feelings, and perceptions. The Stage 1 child recognizes objects as physically separate from the self, but object properties are subject to the child’s perceptions. In this way, objects are not yet psychologically separate from the child’s own experience of them. Kegan indicated that a notable example of what it means to be subject to one’s perceptions is the Piagetian demonstration in which a child is shown a short, wide beaker of liquid and watches as the liquid is poured into a tall thin beaker. Because the level of the liquid is higher in the taller, thinner container, the child perceives there to be more in it, even though he or she was present when the liquid was transferred from one beaker to the other. In short, the child was subject to her immediate perceptions.

Stage 2: Imperial. Although the imperial stage begins in childhood, it often lasts through adolescence and into early college years (Lewis, Bartone, Forsythe, Bullis, Sweeney, & Snook, 2005). Individuals operating from the imperial stage are subject to enduring personal interests, agendas, and role expectations. “Objects” are the perceptions, feelings, and impulses (e.g., the need for immediate gratification) to which one was subject in the previous stage. At Stage 2, one is able to recognize that

individuals (including themselves) possess enduring characteristics (such as honesty, reliability, intelligence, studiousness, etc.). Individuals at Stage 2 are also capable of understanding and considering more than one perspective. However, they are unable to integrate multiple perspectives in such a way that their viewpoints are co-constructed with the perspectives of others (a Stage 3 ability). Instead, they view others' actions and perspectives in light of the *potential consequences* these may have for their own enduring goals or agendas. For example, Lahey et al. (1988) described an interview with a person operating from Stage 2 who discussed being sad about a friend who lied. The upsetting aspect of the situation for this person was that the friend could no longer be counted on for accurate help with an answer if they were studying for a test. In this example, the interviewee was scored as Stage 2 rather than Stage 1 because his statements provided evidence that he was capable of recognizing another's enduring characteristic (such as a tendency to lie). Recognition that the friend's statements may or may not be the truth also revealed an understanding that his own perspective differed from that of his friend. Evidence that the interviewee was not utilizing a framework higher than that of Stage 2 was his inability to view the friend's act of lying beyond its potential consequences for his own goals (such as learning the correct answers for a test). In contrast, a Stage 3 individual might have viewed lying as violating a societal standard of ethical conduct.

Stage 3: Interpersonal. During the interpersonal stage and beyond, individuals continue to have personal goals and agendas, but the goals and agendas are no longer the process by which they organize and make meaningful their experience of the world. Instead, the achievement of personal goals becomes object and the individual becomes subject to a new organizing framework. Kegan proposed that the organizing principles of

Stage 3 are interpersonal connections, shared meaning, and mutual obligations. In the previous stage, individuals were capable of considering viewpoints other than their own; however, they were unable to take these multiple perspectives into account simultaneously. The Stage 3 capacity to consider multiple perspectives simultaneously allows individuals to internalize others' viewpoints in such a way that the self-concept becomes a co-construction of these multiple perspectives. So, for example, my experience of myself (one perspective) simultaneously includes my experience of how I think you view me (another perspective). In addition to internalizing how others think and feel about them in this fashion, individuals at Stage 3 can also internalize the *potential* viewpoints of other persons, organizations, or society, even if these other individuals or institutions are not aware of their actions. As a result, notable differences can be observed in how Stage 3 versus Stage 2 individuals experience guilt. At Stage 3, a guilty party is likely to feel bad for doing something wrong because they can imagine and worry about how another person would now feel about them if they only knew what they had been doing. At Stage 2, feelings of "guilt" are the result of worrying about the possible consequences to one's own interests of getting caught.

Stage 4: Institutional. At Stage 4, individuals construct self-authored systems of values and standards that they use to reflect upon shared meaning. One's self-concept is no longer co-constructed with the opinions of significant others or societal ideals. As a result, Stage 4 individuals enjoy a psychological independence in which they recognize that their values and standards may differ from those of others (or from society). This independence makes it psychologically possible for Stage 4 individuals to comfortably grant others as well as themselves the freedom to possess and apply different standards.

According to Kegan (1994), the challenges of modern adult life are best met through a Stage 4 framework of meaning-making. In the workplace, Kegan indicates that Stage 4 is necessary before employees possess the ability to take ownership of their work because, “to be self-evaluating and self-correcting demands an internal standard,” (Kegan, 1994, p. 169).

Stage 5: Inter-Individual. At Stage 5, individuals are subject to what Kegan refers to as an “interpenetration of systems” (1982). At this stage of meaning-making people become open to considering the truths or value systems of others in a manner that allows them to recognize new values or truths for themselves that they had not allowed themselves to consider before. According to Kegan (1994, p. 311), Stage 4 individuals are able to recognize and “visit” opposing viewpoints like anthropologists, who study and appreciate another culture without judging that culture through the lens of their own value system. A person operating from Stage 5, on the other hand, visits another’s “culture of the mind” open to the possibility that they could be transformed by alternative viewpoints. Stage 4 individuals are identified with their *own* viewpoint. Stage 5 individuals *have* multiple viewpoints and are identified with the universal process of *creating* viewpoints.

Meaning-Making in the Workplace

As stated earlier, Kegan proposed that many of the situations individuals face at work require a Stage 4 or 5 level of meaning-making. However, Stage 3 individuals can also be successful at work, particularly early in their careers. Because Stage 3 is all about mutuality, it is logical to predict that individuals at this level of meaning-making can be excellent team players. Stage 3 managers are able to develop work relationships with

employees based upon “mutual support, promises, expectations, obligations, and rewards” (Kuhnert & Lewis, 1987, p. 652). They are capable of internalizing the goals of the company and typically strive to be viewed in a positive light by supervisors and valued peers. According to Kegan’s description of Stage 3, even when supervisors and peers are not present, potential opinions of significant others are internalized and thereby influence one’s self-concept. Significant influences can include family members, peers, supervisors, or institutions such as the company that employs them, a professional organization, or a church. Because they are “subject to” these shared perspectives, dogmatism is often associated with the Stage 3 framework (e.g., “if you attack my company or religion, then you are also attacking me”).

An individual fully operating at Stage 3 may have no problem working as part of a team in which roles are well defined and members are expected to conform to specific corporate norms. However, Stage 3 employees may experience some internal conflict when supervising subordinates with competing interests or working in situations where expectations are unclear. At this level they are not yet equipped to take ownership of their job positions and cannot comfortably make decisions where it is not possible to please everyone whose opinion and/or respect is important to them.

As managers move into more complex and ambiguous positions, requiring them to rely upon their own judgment rather than guidelines about how to proceed, the self-authored system of values and standards of Stage 4 becomes important for success. Managers at Stage 4 are able to rely upon their personally authored standards to guide behavior and decision-making and are capable of essentially becoming the authors of their jobs (establishing for themselves a set of principles and performance standards).

Kegan (1994) indicated that both Stage 3 and Stage 4 managers are capable of either a “warm and personal” management style, or a more formal, hierarchic, and directive style. However, Kegan also proposed that Stage 3 managers face limitations that include not being able to take a stand if it is unclear what others (including subordinates) want, finding it difficult to say “no,” feeling responsible for the problems of others, or blaming others for things that are actually the manager’s responsibility. By Stage 4, managers are better equipped to empathize with others without internalizing others’ problems. They are able to lead using an internally generated vision, avoid taking responsibility for what is not theirs, and do not attribute responsibility for their own decisions and actions to others.

Other researchers have also proposed that an individual’s stage of meaning-making influences leadership style (Kuhnert & Lewis, 1987) and have demonstrated a relationship between stage level and leadership performance (Bartone, Bullis, Lewis, Forsyth, & Snook, 2001) and military career status (Forsythe, Snook, Lewis, & Bartone, 2002). Researchers have found that managers operating from a higher level of meaning-making approach problem-solving situations by collaborating with others, redefining the problem, and identifying a variety of alternative solutions (e.g., Merron, Fisher, & Torbert, 1987). CEOs operating from a high level of identity development have been reported as more successful at bringing about organizational transformation in a manner that was beneficial to their corporations (Rooke & Torbert, 1998).

At Stage 5, individuals are no longer embedded in their self-established value systems, and are therefore able to recognize and reflect upon the “incompleteness” of these systems – creating the potential for transformation. Although the Stage 4 leader

may feel obligated to communicate a particular vision and bring others “on board,” the Stage 5 leader is more likely to provide a “context in which all interested parties, the leader included, can together create a vision, mission, or purpose they can collectively uphold” (Kegan, 1994, p. 322).

An interesting characteristic of Stage 5 is the manner in which leaders at this level are believed to approach conflict resolution. Kegan (1994) indicated that most contemporary theories and recommended methods of resolving conflict tend to take a Stage 4 approach, which focuses upon developing solutions in a way that considers and respects the differing perspectives involved. However, in the event of protracted disputes (such as those that occur in labor relations or conflicts between nations) a Stage 5 level of meaning-making may be more effective. The Stage 5 approach would involve recognition that the conflicting parties have become too invested in their various viewpoints; Stage 5 negotiators would be better able to focus upon ways to *transform* their opposing views rather than simply attempting to reach a compromise. The current view among constructive-developmental researchers is that few individuals actually evolve to a full Stage 5 level of meaning-making. However, the development of a convenient meaning-making assessment that will provide opportunities for increased research may also help to identify the types of interventions that can promote development to this highest stage.

Meaning-Making and Employee Development. The previous section illustrated how individuals at each meaning-making stage approach their jobs and workplace challenges in a qualitatively different way. An assessment designed to identify an employee’s meaning-making framework could significantly enhance a decision-maker’s

ability to mentor individuals effectively, and appropriately place people in training programs, leadership positions, and job assignments. Consider, for example, the practice of developing employees through challenging job assignments. Developmental job assignments are projects or positions designed to present employees with just enough novelty and challenge to promote learning. These assignments might involve being promoted or transferred to a new position and geographic location, or they can take the form of individual projects or tasks such as managing a group of former peers, dealing with a business crisis, or being assigned an “undoable” project that others have failed to complete in the past (Lombardo & Eichinger, 1989, p. 10). For an assignment to be considered developmental, it must include certain challenging components such as new or broader responsibilities, having to influence people without having authority over them, inheriting problems caused by other employees, supervising difficult employees, or having to negotiate with external groups such as unions or government agencies (Ohlott, Ruderman, & McCauley, 1994).

Although knowledge and experience are important factors when deciding whether an employee is ready for such an assignment, the employee’s meaning-making framework could also be highly relevant. In an assignment that requires a manager to supervise difficult employees, a Stage 2 manager is likely to take a lower-order transactional approach (e.g., promising time off in exchange for overtime), while a Stage 3 leader could be expected to use a higher-order transactional approach in which the “rewards” exchanged for performance are support, trust, and respect (Kuhnert & Lewis, 1984). Are both employees equally prepared to learn from this challenge as well as making a contribution to the organization? Perhaps an employee who is midway through

the transition between Stages 2 and 3 would reap the greatest benefit. If the job challenge requires a Stage 3 meaning-making framework, then it might provide the Stage 2/3 manager with the impetus to complete the transition to Stage 3. This same challenge might be too complex for a Stage 2 manager and not challenging enough for a manager who has already made the transition to Stage 3 or higher. However, if decision-makers do not know where employees are along Kegan's continuum of stages, then these types of assignments might not be granted to the right managers at optimal points in time.

A Final Comment about Subject-Object Interviews versus Quantitative Measures.

Kegan strongly defended the subject-object interview as the best method to obtain a complete understanding of an individual's meaning-making framework (Kegan, Lahey, & Souvaine, 1998). For example, the authors commented that Leovinger's *Sentence Completion Test* (SCT) "yields data that, at best, 'signals' a given stage of development" (p. 40). The interview procedure, on the other hand, allows the researcher to observe mental processes in action and to identify more precisely what is subject versus object for an individual.

Lawrence Kohlberg was similarly skeptical about the usefulness of ratings and rankings to assess individuals' stages of moral reasoning, and used to tease researchers who used the *Defining Issues Test* (DIT), claiming that a multiple-choice test to assess moral reasoning was similar to alchemy (Rest, 1979; Rest, Thoma, Narvaez, & Bebeau, 1997). Kegan and Kohlberg shared a similar viewpoint toward interviews, viewing them as opportunities to observe what was happening in the mind, assuming that individuals were capable of verbalizing their inner processes. Kohlberg claimed that interview scoring was highly valid and relatively error-free (Rest et al., 1999a). Rest (1979)

pointed out that, despite the rich information that can be obtained from interviews, the interview methodology had problems. For example, interviewees who were not very articulate might be scored at a lower stage because they were unable to elaborate verbally on their reasoning. Another concern with the Kohlberg interview was that participants were asked to choose a course of action for a hypothetical moral dilemma and then defend that course of action through post hoc reasoning. Rest's (1979) concern with post hoc reasoning was that the researcher could not know "whether the subject's reasons had influenced his original decision or whether the procedure is forcing the subject to invent rationalizations for a previous commitment" (p. 88).

Despite subjectivity, the subject-object interview will probably remain the most effective method for gaining a complete understanding of an individual's framework for meaning-making. Interviewers can probe for information about how a person makes-meaning, as well as identifying a person's limits, the point beyond which they cannot go (Lahey et al., 1988). However, there are potential advantages to having a measurement scale that can reliably *estimate* developmental stage level. For much research, where interviews are too resource intensive, such a measure would facilitate studies on relationships between levels of meaning-making and other variables of interest in a variety of settings. The applied advantage is to have an efficient and cost-effective method of at least approximating the stage level of individuals in the workplace in order to lead and develop employees effectively.

Development of RED

Because the purpose of this research was to test a new measure, this section of the paper provides a description of RED and how the assessment was developed. RED was

designed to identify how respondents would react to each of five hypothetical dilemmas. Participants were scored by the types of issues they identified as most significant or relevant to them in each dilemma.

Development of the Dilemmas. The first step in the development RED was to create hypothetical dilemmas that were likely to elicit reactions revealing an individual's subject-object balance. An important issue to resolve at this point was the nature of the dilemmas that would be presented. Should the stories describe extraordinary dilemmas, similar to those that Kohlberg (1969, 1981, 1984) developed to assess moral reasoning, or should they describe situations that are more commonplace? Kegan (1994) proposed that less than one-half of adults have developed the capacity to manage the complexity of their workplace and personal lives effectively. What is fundamental about Kegan's proposition to the development of a dilemma-based assessment is that the circumstances that present these demands are not necessarily extraordinary. On the contrary, situations that can lead to the "in over our heads" phenomenon (Kegan, 1994) are ever-present at work, school, and home. The implication for our work lives is that our developmental progress can influence our reactions and effectiveness in ordinary predicaments such as dealing with a team member who is not contributing a fair share, coping with organizational change, managing competing constituencies, negotiating office politics, and resolving co-worker conflicts. These are "everyday dilemmas."

The RED assessment attempted to elicit reactions to situations that were challenging, while at the same time being so familiar that the respondents would be likely to identify with the underlying themes. Before writing the dilemmas, the researcher reviewed real life problems that had been submitted to workplace and general advice

columns and identified several common themes that laid the groundwork for the vignettes. These themes included balancing personal versus professional relationships between supervisors and employees in the workplace, conflicts among employees, high-level managers who cannot get along, competition among co-workers, and dealing with organizational change. Story ideas were also obtained from an earlier study in which students provided written narratives about personal emotional experiences (Bellenger, 1999).

The information sources described above provided ideas for 14 stories about workplace and non-workplace dilemmas. To identify realistic reactions to the 14 dilemmas, 25 undergraduate psychology students were asked to review the stories and provide feedback. During a 45-minute interview, each student read three or four stories and discussed how they would have reacted if they had experienced the described situations. The students' comments were useful for developing reaction statements that represented Kegan's Stages 2 and 3. Most undergraduate students advance from Stage 2 to Stage 3 during their college years (Lewis et al., 2005). The development of Stage 4 and Stage 5 reaction statements relied upon Kegan's published descriptions of what individuals in these higher stages are capable of expressing, as well as transcripts from past subject-object interviews in which respondents were scored at Stage 4 or higher.

The students who reviewed the original 14 dilemmas provided suggestions on ways to clarify the dilemmas and make them more engaging. There was one recommendation that globally affected all of the dilemmas. Originally, most of them were written in the third person, and readers had to imagine themselves in a main character's position. It was suggested that the dilemmas would be more engaging if they

were written in the second person and immediately asked readers to imagine themselves in a particular situation. Following this suggestion, all dilemmas were revised so that they spoke to readers using second person pronouns and all began with the word “imagine.”

First Version of RED. Originally, RED was designed to include all 14 dilemmas (or possibly seven different dilemmas for two versions of the test). Approximately 32 reaction statements were written for each dilemma, and each statement was written to reflect a particular Kegan stage. The reaction statements also included “high sounding” items (statements that sound sophisticated, but are essentially meaningless). Similar nonsensical statements have been included on the *Defining Issues Test* to identify respondents who were motivated to sound sophisticated but were not attending to the meaning of each statement (Rest et al., 1997).

After completing the first version of RED, the items were reviewed in a random order (not stage related). It was concluded that significant revisions were required, because the response items were too brief to represent specific stages of meaning-making. Another problem was that the test was too long, even when split into two seven-scenario versions. A subsequent round of revisions produced a shorter instrument with fewer reaction statements that were written to provide more depth and stage-level differentiation (five dilemmas with 13 reaction statements each).

Second Version of RED. The version of RED tested in the present study consisted of 5 dilemmas selected from the original 14. Three of the five stories depicted work-related situations: (a) being promoted to a new job position that required the supervision of former co-workers; (b) recommending a good friend for a vacant position within their

organization, only to have him/her outperform them at work; and (c) working as part of a team that had not fared well within the company because the team's manager did not get along well with his/her boss. The remaining two dilemmas were (d) not being able to get along with a significant other's good friend; and (e) chairing a project committee for a charitable organization and being criticized by an experienced committee member who did not agree with one's ideas.

After reading each dilemma, respondents were asked to write an open-ended appraisal of the situation. In writing their appraisal, respondents were asked to consider how they would feel in the described situation and why they would feel that way. The purpose of this exercise was to help participants become engaged in each dilemma, and to provide ideas for further revisions of the measure.

In addition to encouraging participants to become engaged with each dilemma, the open-ended question was also intended to enhance self-focus (also referred to as private self-awareness) before respondents rated the reaction statements for each dilemma. Private self-awareness has been defined as "awareness of oneself from a personal perspective" (Fejfar & Hoyle, 2000, p.132) and methodologies for increasing private self-awareness include exposing respondents to a mirror, having participants listen to their own voices or having them write stories about themselves. Increasing private self-awareness can have either positive or negative effects, depending upon the situation and variables of interest. For example, researchers have demonstrated that private self-awareness can increase helping behavior, while simultaneously increasing cheating behavior (Malcolm & Ng, 1989). Private self-awareness can also increase

negative thoughts or pessimism because it involves self-evaluation (Pyszczynski, Hamilton, & Herring, 1989; Pyszczynski, Holt, & Greenberg, 1987).

Self-focus is relevant to the present study because there is evidence that it improves the internal consistency of personality scales (Hamilton & Shuminsky, 1990). It has been demonstrated in the past that items that appear later in a measure have greater item-total correlations, and this result has been explained as the result of early items serving to activate self-schemas (Knowles, 1988). Having a preceding activity that increases self-focus is believed to have a nullifying effect on item serial position.

Several studies have successfully utilized the personal story telling method of increasing self-focus developed by Fenigstein and Levine (1984) in which respondents are instructed to include words such as “I” or “myself” in a narrative (e.g., Hamilton & Shuminsky, 1990; Pyszczynski et al., 1989; Pyszczynski et al., 1987). In the present study, the purpose of having respondents describe how they would react to each dilemma was to increase self-focus, thereby activating a schema that would hopefully enhance the internal consistency of the measure.

Another potential benefit of the narrative responses was that participants might provide information that could be used for future improvements of RED items. One final benefit of the open-ended question was to assess the extent to which participants were motivated to respond thoughtfully to the assessment.

The open-ended appraisal was followed by three sets of reaction statements to be rank ordered. Each set included statements representative of Stages 2, 3, 4, and 5, and respondents were asked to rank order the statements within each set on the basis of the following criteria: each statement’s significance as an underlying issue in the dilemma

(set 1) and the likelihood that they might react in a manner described by each statement (sets 2 and 3). Rankings rather than ratings were collected in order to force differentiation among the statements in each set.

In the first ranking question, the statements described potential *underlying issues* that could be attributed to the scenario. Respondents were asked to rank order these issues from *most significant* (1) to *least significant* (4). The second ranking question presented statements that described potential social or interpersonal reactions to the scenario. Respondents were asked to consider how they would feel if they were in this situation, and then rank order the statements from their *most likely reaction* (1) to *least likely reaction* (4).

The third ranking question presented statements that reflected intrapersonal issues, focusing upon the self-reflection that each dilemma might produce. Once again respondents were asked to rank order the statements from their *most likely* (1) to *least likely reaction* (5). A fifth statement was included in the intrapersonal group for each dilemma. This fifth item was not a stage-related statement. Instead, it was a “high sounding” item – a statement that sounded complex but was, in reality, nonsensical. Frequent endorsement, or high rankings, of high sounding items can indicate that the test-taker is either not paying attention to the content of the statements and/or trying to present oneself in a socially desirable fashion. Such items have been used successfully within the *Defining Issues Test* (Rest et al., 1999) to identify invalid cases.

In summary, the version of RED tested in Study 1 was structured as follows: (a) participants were asked to read five stories describing relatively common workplace or personal dilemmas, (b) each story was followed by an open-ended question requesting

reactions to the situation, and (c) the open-ended question was followed by three sets of statements to be rank-ordered by respondents.

Expert Ratings

Two researchers who were experienced in conducting and scoring subject-object interviews reviewed RED's reaction statements in a random order and categorized each statement as being most representative of Stage 2, 3, 4, 5 or a nonsense item. Raters were instructed to place a question mark next to any statement that did not appear to be typical of one particular meaning-making stage. Rater 1 had participated in reaction statement development but had not reviewed the items for several months. Rater 2 had never seen the statements before performing this task. Rater instructions are shown in *Appendix A*.

Rater agreement was 94%. Out of 65 reaction statements Rater 1 placed all statements into the categories that they were intended to represent. Rater 2 incorrectly categorized 4 items by labeling 2 Stage 5 items as Stage 4 and vice-versa.

Overview of the Research

The present research tested the psychometric properties and construct validity of the *Reactions to Everyday Dilemmas* (RED) instrument. Convergent validity was assessed by comparing scores on RED to scores of moral development obtained from the *Defining Issues Test* (Rest, 1975, 1979; Rest, et al., 1999b). To establish discriminant validity, RED was compared to scores of dispositional optimism (Scheier & Carver, 1985; Scheier, Carver, & Bridges, 1994). The sections that follow provide descriptions of each construct validity measure and the rationale for including them in the present study.

Moral Reasoning and the Defining Issues Test (DIT). Lawrence Kohlberg's theory of moral development states that, over the lifespan, moral reasoning evolves both qualitatively as well as in terms of increased complexity. Kohlberg (1984) described six stages, which he grouped into three levels: pre-conventional, conventional, and post-conventional. Kegan, a student of Kohlberg, viewed moral development as one facet of the meaning-making progression throughout the lifespan and proposed that his developmental theory identified the "underlying deep structure" that drives change in moral reasoning (Lahey, 1986, p. 14). Kegan stated that, "each of Kohlberg's stages, like each of Piaget's, may be the consequence of a single underlying process of evolution" (Kegan, 1982, p. 71).

Kegan (1982) indicated that the subject-object balance during Kohlberg's Stages 2, 3, and 4 corresponded with his meaning-making stages 2, 3, and 4 respectively, and that DIT Stages 5 and 6 corresponded with Kegan's Stage 5. However, a case can be made for the proposition that Kohlberg's Stages 2 and 3 roughly correspond to Kegan's Stage 2, that Kohlberg's Stage 4 corresponds to Kegan's Stage 3, Kohlberg's Stage 5 corresponds to Kegan's Stage 4, and that only Kohlberg's Stage 6 corresponds roughly with Kegan's Stage 5 (P. Lewis, personal communication, September 17, 2006). Brief descriptions of Kohlberg's moral reasoning stages are provided below.

In the pre-conventional level, Stage 1 perceptions of moral or ethical behavior are driven by the desire to avoid the negative consequences associated with breaking rules established by higher authorities (parents, teachers, etc.). In Stage 2, one's reasoning expands to a consideration of fairness (e.g., the legitimacy of acting in one's own best interest, recognizing that others have a right to act in their own best interests as well).

Kohlberg indicated that the concept of equity is defined in terms of an individual's needs rather than intent. For example, "it can be fair for the poor to steal because they *need* the food." (1984, p. 627). Although individuals at Kohlberg's Stage 2 accept that they or others may have to break laws in order to meet their needs, they are unable at this point to integrate conflicting perspectives at one time, similar to the limitations of Kegan's Stage 2. Therefore, in Kohlberg's dilemma about a man who could not afford to purchase a cancer drug for his dying wife (Heinz), a Stage 2 respondent might claim on one hand that it is fair for Heinz to steal the drug because he needs it, and then later indicate that "the judge should punish Heinz, because if he doesn't then others may try to get away with stealing." (Kohlberg, 1984, p. 627).

Once individuals move into the conventional level of Stage 3, moral reasoning develops to encompass the social expectations of important people or institutions. On the surface, descriptions of Kohlberg's Stage 3 resemble Kegan's Stage 3, because the Stage 3 moral perspective has expanded to include consideration of how individuals should behave in "mutually trusting relationships" (p. 628). However, Kohlberg's Stage 3 remains a very transactional stage, in which morality is viewed from a perspective of reciprocity or what Kohlberg referred to as "golden rule role-taking" (p. 629). Although relationship maintenance becomes important in Kohlberg's Stage 3 (resembling Kegan's Stage 3), concepts such as mutual trust and loyalty are defined by an expectation of reciprocity (Kegan's Stage 2).

At Kohlberg's Stage 4, individuals develop the capacity to consider the importance of society and their place in it. There is the perception that laws have been created for the common good of all, and should therefore be obeyed even if one does not

agree with them. Kohlberg's fourth stage of moral reasoning roughly parallels Kegan's Stage 3. Kohlberg's Stage 4 and Kegan's Stage 3 are influenced significantly by societal expectations and doing what is best for the group. Individuals operating at Kegan's Stage 3 are influenced by individual relationships in a manner similar to the societal influence that is characteristic of Kohlberg's Stage 4. Thus, individuals are likely to feel that certain behaviors are important to maintain love, trust, or respect in relationships, simply for the good of the relationship.

At Kohlberg's Stage 5, there is recognition of and appreciation for universal rights that transcend the laws and rules that may have been established by society or by one's own subgroup and these rights must be upheld even if they violate society's laws. The meaning-making framework of Kegan's Stage 4 is similar to the moral reasoning framework of Kohlberg's Stage 5 because both require the capacity to develop a value system that transcends established rules, procedures, laws, etc.

At Stage 6, Kohlberg proposed that individuals believed moral decisions should be approached as though they did not know which side or position they would occupy, and decisions should be made to provide all involved parties with an adequate solution. At Kohlberg's sixth stage of moral reasoning and Kegan's Stage 5, individuals possess the capacity to hold multiple perspectives and are no longer defined by their personal value systems.

Pratt, Diessner, Hunsberger, Pancer, and Savoy (1991) reported a significant correlation between subject-object scores and the weighted average score (WAS) from Kohlberg's Moral Judgment Interview, providing preliminary empirical support for an hypothesized correlation between subject-object scores and moral reasoning stages.

Given the correlation between Kohlberg's interview assessment and the DIT (Rest, 1979), it was anticipated that scores on RED would also positively correlate with scores on the DIT measure of moral reasoning. This expectation led to the first research hypothesis of the present study.

H1: Stage-level scores from RED will positively correlate with an individual's stage of moral reasoning as measured by the *Defining Issues Test* (Rest, 1975, 1979; Rest, Narvaez, Bebeau, & Thoma, 1999b).

Similar to the structure of RED, the DIT presents respondents with a series of dilemmas. However, the DIT dilemmas are designed to activate a moral schema, while RED's dilemmas describe problems that do not call for a moral judgment. Numerous studies have demonstrated the construct validity of the DIT over the past 25 years (Rest, Thoma, & Edwards, 1997).

Dispositional Optimism as a Measure of Divergent Validity. For the past two decades, research has suggested that dispositional optimism (Scheier & Carver, 1985) has a beneficial impact upon one's ability to deal with a variety of stressful situations. For example, studies have demonstrated that optimistic men and women cope better with health problems (Affleck, Tennen, Zautra, Urrows, Abeles, & Karoly, 2001; Steginga, & Occhipinti, 2006); recover more quickly after surgery (Scheier et al., 1989); and more easily adapt to college life (Brissette, Scheier, & Carver, 2002).

It has been suggested that differences between optimists and pessimists lie in their expectancies for the future, which then influence the strategies they employ to deal with stressors (Scheier, Weintraub, & Carver, 1986; Scheier, Carver, & Bridges, 1994).

Optimists use problem-focused coping strategies or, when faced with irresolvable problems, are more likely to demonstrate acceptance, humor, or positive reinterpretations of the situation. Pessimists, on the other hand, are more likely to use a strategy of denial, even if there is something that can be done to solve the problem (Scheier, et al., 1986, 1994).

Dispositional optimism was assessed in the present research with the Revised *Life Orientation Test* (Scheier, et al., 1994). Research using the original *Life Orientation Test* (LOT) and the 1994 revised version (LOT-R) has demonstrated moderate correlations with several personality characteristics such as self-mastery (+), trait-state anxiety (-), self-esteem (+), and neuroticism (-) (Sheier et al., 1994). Relevant to the present study, researchers have *not* suggested that there is any relationship between dispositional optimism and age, educational level, or, most significantly, adult development. Optimists from a wide range of age groups have been studied, from college freshmen to men and women in their fifties and older, and there has been no indication of a developmental component to this dimension of personality. Therefore, scores of dispositional optimism, as measured by the LOT-R, were not expected to correlate with stage-level scores from RED. This led to the second research hypothesis.

H2: Stage-level scores from RED will not correlate with an individual's level of dispositional optimism as measured by the Revised *Life Orientation Test*.

In summary, the successful validation of RED would represent a first step in the development of a quantitative meaning-making assessment. In the workplace, the convenience of this measure would give employers a useful assessment tool for employee

development. To develop and validate the measure, two studies were conducted. In Study 1, a pilot test was conducted to test the experimental measure for scale reliability and correlation with respondents' ages and educational levels. For Study 2, a modified version of RED was administered along with measures to establish convergent and divergent validity.

II. STUDY 1

Method

Participants. Study 1 involved 183 participants. The sample included 159 undergraduate students who participated for extra credit in a psychology or management course, 4 graduate students and 21 employees of a local hospital. Respondents ranged in age from 19 to 56, with a mean age of 22.39 years ($SD = 6.81$) and a median age of 20 years. Fourteen participants were removed from the analysis for submitting incomplete data, resulting in a sample of 169. Table 1 presents the sample characteristics of gender, highest educational level achieved, and college classification.

Table 1
Summary of Study 1 Sample Characteristics

Sample Characteristics	Percentage ($N = 169$)
Gender	
Men	33.7
Women	65.7
No response	0.6
Educational level	
High school diploma or GED	75.7
2-year degree or technical school degree	11.2
4-year degree	8.3
Professional Program Degree	0.6
Master's degree or doctorate	2.4
Other	1.2
No response	0.6
College classification	
Not currently enrolled in college	8.3
Freshman	27.8
Sophomore	17.2
Junior	24.9
Senior	18.9
Master's or doctoral student	2.5
Other	0.6

Materials. Participants completed the RED measure and a set of demographic questions. Questionnaire instructions and a complete set of questions for one of the five dilemmas are shown in *Appendix B*. At the end of the assessment, respondents were invited to provide written feedback about the dilemmas. Participants were also asked to indicate whether the task of writing open-ended appraisals of each dilemma helped or hindered their ability to complete the subsequent statement rankings.

Procedure. Student participants were recruited through flyers and e-mailed announcements in the Psychology, Counseling and Management Departments of a large southern university. Students attended group data-collection sessions that lasted approximately 45 minutes. For hospital employees, the researcher gave 30 measures to two shift supervisors to distribute to clinicians on their shifts. Each envelope contained the measure, an information sheet and instructions to complete the test at home and return it to a designated drop box. Hospital employees and graduate students were given a small gift for their participation, while undergraduate students received extra credit in a psychology or management course.

All respondents were assured that their participation was voluntary and that their responses would be anonymous. To protect the anonymity of hospital employees who participated, gifts for participation were left sitting out by the drop box for participants to take on the honor system. Hospital employees returned 21 valid questionnaires.

Results

Respondent rankings were reverse coded so that a high score represented high endorsement of a statement. Rankings for the first two sets of items ranged from four to

one, and from five to one on the third set in which a “high sounding” item was added.

Table 2 shows the average rankings for items within each story.

An assessment of internal consistency was performed for each stage scale. Scales for each of the four Kegan stages were comprised of 15 items (three per story) that were written to represent prototypical issues, interpersonal reactions or intrapersonal reactions of individuals at each stage. Internal consistency was assessed by calculating Cronbach’s alpha and by examining item-total correlations. Three reliability analyses were conducted for each stage scale. The first reliability analysis was performed for the total sample, without removing respondents who gave high rankings to the nonsensical high-sounding items. The second analysis explored differences in alpha coefficients when respondents were screened out for selecting one or more high-sounding items as their highest-ranking choice within a set. The final analysis attempted to improve internal consistency by removing items that had extremely low or negative correlations with the scale mean. Tables 3 through 6 present item statistics for each of the four stage-level scales. Before screening unreliable respondents from the sample, Cronbach’s alpha coefficients for Stages 2, 3, 4 and 5 were .50, .35, .38, and .46 respectively.

Analysis of Respondent Screening. To explore the impact on reliability of respondents who might not have paid close attention to the items, a reliability analysis was performed after removing cases in which participants had selected at least one high sounding item as their top choice within a set. Seventy-nine percent of the sample (133 respondents) completed the measure without ranking a high-sounding item as a most significant issue or most likely reaction. When the 36 respondents who preferred at least one high-sounding item were removed from the analysis, the alpha coefficients rose a few

points for two of the stages and fell for the other two (α 's = .55, .32, .31, and .49 for Stages 2 to 5 respectively).

Another reliability analysis was performed among respondents who preferred no more than one of the high-sounding items. Of the 36 participants who endorsed high-sounding items, 27 only did so once. The remaining nine participants either selected two or three of the nonsense items as their preferred choice, and these nine were excluded from the new analysis. Alpha coefficients and item statistics for the analysis that excludes nine respondents are presented in Tables 3 through 6 along with the original total sample statistics. This less stringent filter resulted in reliability estimates of .53, .35, .37, and .48 for Stages 2 through 5, which slightly improved the reliability of the scales for Stage 2 and 5 while lowering reliability one point for the Stage 4 scale.

Item Analysis. The item statistics presented in Tables 3 through 6 are ranked ordered by their item-total correlations for the total sample (from worst to best). The tables also show how the alpha coefficient would change if an item were deleted. To improve scale reliability, items with the worst item-total correlations were removed, one at a time, until a point was reached where the removal of additional items resulted in a decrease in reliability.

Table 7 shows the improved scale reliability estimates for each scale after removing four to five items and screening out participants who highly endorsed at least two high-sounding items. Although the reliability of each scale improved, the alpha levels remained low (from .45 to .63).

Table 2
Study 1 Mean Ranking for Each Item by Story in the RED Measure

Item	Story A	Story B	Story C	Story D	Story E
Stage 2 Issue	3.34 (.97)	3.24 (1.00)	2.20 (1.20)	1.77 (1.03)	2.15 (1.17)
Stage 2 Inter	2.95 (.98)	2.33 (.90)	2.12 (.99)	2.36 (1.04)	2.62 (1.05)
Stage 2 Intra	3.15 (1.40)	2.58 (1.27)	2.33 (1.11)	3.27 (1.23)	3.53 (1.44)
Stage 3 Issue	2.56 (1.02)	2.49 (.98)	2.62 (1.06)	2.80 (.93)	2.41 (.95)
Stage 3 Inter	1.84 (1.01)	2.10 (1.03)	1.98 (.89)	2.25 (1.14)	2.28 (1.09)
Stage 3 Intra	2.62 (1.39)	3.54 (1.19)	3.93 (1.12)	4.21 (1.12)	3.27 (1.31)
Stage 4 Issue	1.98 (.81)	2.28 (1.11)	2.83 (1.06)	3.31 (.95)	3.21 (1.03)
Stage 4 Inter	2.94 (1.12)	3.53 (.82)	3.53 (.87)	2.70 (1.03)	3.02 (1.00)
Stage 4 Intra	3.60 (1.27)	4.22 (1.01)	3.75 (1.15)	2.42 (1.26)	3.56 (1.31)
Stage 5 Issue	2.12 (1.11)	2.00 (.99)	2.36 (1.05)	2.12 (.87)	2.22 (1.00)
Stage 5 Inter	2.28 (.95)	2.04 (1.01)	2.37 (.99)	2.67 (1.20)	2.08 (1.10)
Stage 5 Intra	3.07 (1.41)	2.47 (1.32)	3.48 (1.07)	2.63 (1.28)	2.59 (1.21)
High Sounding	2.57 (1.35)	2.20 (1.16)	1.51 (.87)	2.50 (1.33)	2.06 (1.17)

Note. Standard deviations are shown in parentheses. $N=169$.

Table 3
Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 2 Scale

Stage 2 Items	Total Sample, 15-Item Stage 2 Scale ($N = 169, \alpha = .50$)		Sample Endorsing < 2 High Sounding Items, 15-Item Stage 2 Scale ($N = 160, \alpha = .53$)	
	Corrected Item-Total Correlation	Alpha if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Story A: Stage 2 Inter	-.1550	.5478	-.1400	.5682
Story D: Stage 2 Intra	.0078	.5264	.0236	.5484
Story B: Stage 2 Inter	.0178	.5147	.0270	.5388
Story A: Stage 2 Issue	.0180	.5164	.0218	.5409
Story C: Stage 2 Intra	.0592	.5115	.0601	.5375
Story B: Stage 2 Issue	.1193	.4977	.1279	.5229
Story E: Stage 2 Intra	.1971	.4818	.2164	.5054
Story C: Stage 2 Inter	.2052	.4807	.2121	.5072
Story C: Stage 2 Issue	.2303	.4730	.2660	.4937
Story B: Stage 2 Intra	.2720	.4615	.2911	.4866
Story E: Stage 2 Inter	.2944	.4609	.2954	.4900
Story D: Stage 2 Issue	.3083	.4586	.3134	.4865
Story A: Stage 2 Intra	.3239	.4444	.3277	.4746
Story E: Stage 2 Issue	.3464	.4451	.3747	.4699
Story D: Stage 2 Inter	.3730	.4439	.3965	.4701

Table 4
Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 3 Scale

Stage 3 Items	Total Sample, 15-Item Stage 3 Scale		Sample Endorsing < 2 High Sounding Items, 15-Item Stage 3 Scale	
	<i>(N = 169, $\alpha = .35$)</i>		<i>(N = 160, $\alpha = .35$)</i>	
	Corrected Item-Total Correlation	Alpha if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Story E: Stage 3 Intra	-.0443	.3961	-.0613	.4015
Story E: Stage 3 Inter	.0238	.3649	.0109	.3685
Story C: Stage 3 Intra	.0518	.3567	.0330	.3610
Story B: Stage 3 Inter	.0684	.3503	.0450	.3563
Story B: Stage 3 Issue	.0691	.3496	.0652	.3500
Story E: Stage 3 Issue	.0750	.3478	.0847	.3444
Story B: Stage 3 Intra	.1056	.3395	.1073	.3381
Story D: Stage 3 Intra	.1153	.3361	.1067	.3381
Story C: Stage 3 Inter	.1188	.3362	.1629	.3239
Story A: Stage 3 Intra	.1261	.3327	.1377	.3270
Story A: Stage 3 Inter	.1414	.3286	.1501	.3254
Story D: Stage 3 Issue	.2012	.3133	.2043	.3117
Story C: Stage 3 Issue	.2043	.3079	.2059	.3063
Story D: Stage 3 Inter	.2097	.3038	.2099	.3025
Story A: Stage 3 Issue	.2423	.2973	.2551	.2926

Table 5
Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 4 Scale

Stage 4 Items	Total Sample, 15-Item Stage 4 Scale (<i>N</i> = 169, α = .38)		Sample Endorsing < 2 High Sounding Items, 15-Item Stage 4 Scale (<i>N</i> = 160, α = .37)	
	Corrected Item-Total Correlation	Alpha if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Story A: Stage 4 Inter	-.1301	.4421	-.1236	.4269
Story D: Stage 4 Intra	-.0985	.4417	-.1410	.4419
Story A: Stage 4 Issue	-.0929	.4147	-.0769	.3965
Story D: Stage 4 Inter	.0496	.3870	.0924	.3580
Story B: Stage 4 Issue	.0819	.3788	.0807	.3625
Story E: Stage 4 Inter	.1580	.3562	.1535	.3401
Story B: Stage 4 Intra	.1630	.3546	.1295	.3471
Story C: Stage 4 Intra	.1745	.3491	.1509	.3390
Story A: Stage 4 Intra	.2035	.3369	.1894	.3239
Story C: Stage 4 Issue	.2103	.3394	.1829	.3300
Story B: Stage 4 Inter	.2189	.3450	.2573	.3171
Story E: Stage 4 Intra	.2315	.3252	.2460	.3003
Story C: Stage 4 Inter	.2430	.3367	.2053	.3311
Story E: Stage 4 Issue	.2469	.3291	.2533	.3087
Story D: Stage 4 Issue	.2488	.3320	.2464	.3141

Table 6
Study 1 Reliability Analysis and Item-Total Statistics for the RED Stage 5 Scale

Stage 5 Items	Total Sample, 15-Item Stage 5 Scale ($N = 169, \alpha = .46$)		Sample Endorsing < 2 High Sounding Items, 15-Item Stage 5 Scale ($N = 160, \alpha = .48$)	
	Corrected Item-Total Correlation	Alpha if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Story B: Stage 5 Inter	-.0332	.4883	-.0460	.5050
Story B: Stage 5 Issue	.0021	.4803	.0289	.4895
Story E: Stage 5 Inter	.0875	.4637	.0795	.4802
Story A: Stage 5 Intra	.1071	.4640	.1015	.4813
Story C: Stage 5 Intra	.1362	.4517	.1501	.4640
Story D: Stage 5 Issue	.1479	.4495	.1410	.4662
Story E: Stage 5 Intra	.1513	.4484	.1420	.4666
Story A: Stage 5 Issue	.1542	.4474	.1524	.4635
Story D: Stage 5 Intra	.1619	.4458	.1647	.4609
Story A: Stage 5 Inter	.1622	.4461	.1744	.4590
Story C: Stage 5 Issue	.1624	.4455	.2019	.4523
Story C: Stage 5 Inter	.1901	.4397	.1962	.4542
Story B: Stage 5 Intra	.2074	.4323	.2206	.4448
Story E: Stage 5 Issue	.3518	.4020	.3748	.4134
Story D: Stage 5 Inter	.3821	.3832	.3962	.3966

Table 7
Study 1 Alpha Coefficients after Removing Cases and Items for the RED Measure

Scale	Total Sample, 15-Item Scales (<i>N</i> = 169)	Number of Items Deleted	Sample Endorsing < 2 High Sounding Items, Scales with Low Performing Items Removed (<i>N</i> = 160)
Stage 2	.50	5	.63
Stage 3	.36	4	.45
Stage 4	.37	5	.58
Stage 5	.48	4	.53

Correlation Analysis. Mean scores were calculated for each of the reduced stage scales for a correlation analysis with age, education, and college classification.

Educational levels and college classifications were coded in an ordinal manner so that more advanced levels were assigned higher numeric codes. Table 8 shows significant positive correlations between age and scores on the higher-level stage scales 4 and 5, $r_s = .227, .297$ respectively, $p < .01$. Scores from the Stage 5 scale also correlated positively with education ($r_s = .247, p < .01$) and college classification ($r_s = .182, p < .05$).

Significant negative correlations were observed between scores on the lower level Stage 2 scale with age ($r_s = -.218, p < .01$), education ($r_s = -.259, p < .01$) and college classification ($r_s = -.191, p < .05$). Stage 3 scores also correlated negatively with age and educational level ($r_s = -.203$ and $-.171$, respectively, $p < .05$).

Table 8
Study 1 Bivariate Correlations among the RED Stage Scales, Age, Education, and College Class

	Stage 2	Stage 3	Stage 4	Stage 5	Age	Education	College Class
Stage 2	.63	.095	-.500**	-.576**	-.218**	-.252**	-.191*
Stage 3		.45	-.271**	-.364**	-.203*	-.172*	-.128
Stage 4			.58	.204**	.227**	.134	.124
Stage 5				.53	.295**	.229**	.182*
Age					---	.601**	.852**
Education						---	.417**
College Class							---

Note. Correlation coefficients were calculated using the Spearman rho procedure for ordinal data. Cronbach alpha coefficients are shown in the diagonal. Sample sizes ranged from $N = 145$ to $N = 160$.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

To test the relationship between age, educational level and RED scores, partial correlations were calculated between each stage score and age while controlling for educational level. With the education control, the correlation coefficients became non-significant for Stages 2 and 3 and less significant for Stages 4 and 5. The partial correlation coefficients for age and Stages 2, 3, 4, and 5, respectively were $-.088$ ($p = .259$), $-.114$ ($p = .145$), $.205$ ($p = .008$), and $.173$ ($p = .026$).

Intercorrelations among the stage scale scores were compatible with the theoretical order of Kegan's stages. Stage 2 scores correlated negatively with scores for Stage 4 ($r_s = -.500, p < .01$) and Stage 5 ($r_s = -.576, p < .01$). Stage 3 scores correlated negatively with Stages 4 and 5 ($r_s = -.271$ and $-.364$ respectively, $p < .01$), and Stage 4 correlated positively with Stage 5 ($r_s = .204, p < .01$).

Study 1 Discussion

Mean stage-level scores appeared to be related to one another, age and education in a manner that would be expected from a personality measure that identifies stages of development. Negative correlations are inherent to ipsative scales such as the RED ranking scales (Alwin & Krosnic, 1985; Krosnick & Alwin, 1988). However, the pattern and strength of the negative and positive scale intercorrelations were theoretically consistent, as the negative intercorrelations became stronger for pairs of stages that were further apart. Age and educational level correlated negatively with scores on Stages 2 and 3, and correlated positively with scores for the advanced stages (4 and 5). The finding that the age correlations weakened or disappeared after controlling for educational level is a positive sign that RED scores respond to statistical manipulations in a manner consistent with other indices of adult development such as the subject-object interview and the moral judgment interview (Pratt et al., 1991). (It should be noted that Pratt et al. observed modest negative correlations between age and scores of adult development that became non-significant when they controlled for educational level.)

Despite the significant correlations and their patterns, the low reliability coefficients for each of the scales were a problem. None of the scales achieved Nunnally's (1978) minimally accepted alpha level of .70, even after removing four to five of the poorer items. With 15 items per scale, length was not a factor that contributed to low reliability. Instead, the items simply had low inter-item and item-total correlations. There are several factors that probably influenced scale reliability.

The first issue was the sample, which undoubtedly had an impact on reliability. Most Study 1 participants were college undergraduates and the median age of the sample

was 20 years. Because the construct being measured was related to age, a greater diversity of ages might have improved reliability. However, the decision was made to move ahead with Study 2 and a revised version of the measure rather than continuing to collect Study 1 data, because there were two other issues that needed to be addressed: (a) scaling and (b) participant instructions.

Scaling. The use of rankings versus ratings has been somewhat controversial. For ipsative data, issues include a debate about the accuracy of reliability estimates and the appropriateness of certain statistical procedures (Bartram, 1996; Saville & Willson, 1991), and an increased burden upon respondents who find it more difficult to rank items as compared to assigning ratings (Alwin & Krosnick, 1985). Normative data (ratings) are easier for respondents to generate. However, one criticism is that this “ease” of providing ratings may result from subjects not expending energy to differentiate among items. In addition, rating scales are subject to response bias from raters who favor certain parts of the scale (e.g., highly endorsing all items or using only the center of a scale) (Saville & Willson, 1991). Ipsative scales force respondents to differentiate between items and prevent them from using only one section of a scale.

Despite the item differentiation and reduction of response bias mentioned above, the use of ipsative scales (ranked items) in the Study 1 version of RED provided a limited amount of information on which to base future decisions about item revisions or permanent deletions. In most of the studies mentioned above, researchers utilized both types of scales (ratings and rankings) to evaluate the relative performance of each format. The inclusion of normative scales in the RED measure would provide additional item information and an estimate of reliability for normative scales. When combined with

rankings, ratings could help respondents make ranking decisions. In the DIT, participants refer to their ratings to when deciding how to rank items. This combination of ratings and rankings also provides a consistency check to determine whether or not test takers are actually paying attention. Therefore, for Study 2, a revised version of RED was created with both ratings and rankings.

Participant Instructions. An issue that arose with the Study 1 version of RED was that the limited number of items in each set (four or five) could not include the range of possible reactions that an individual might experience in each story. Several participants expressed concern to the researcher (both verbally and in writing) that none of the response items described *exactly* how they would have reacted. This finding may have resulted in some rating sets being left blank or completed without much thought. To address this issue, new instructions were added to the Study 2 version of RED emphasizing that the reaction statements would not necessarily include their most likely reactions.

III. STUDY 2

Method

Participants. Study 2 included 155 participants. The sample included 80 undergraduate students who participated for extra credit in a psychology course, 9 public school system employees, and 66 employees of a local hospital. Of the 155 participants, 22 were removed from the analysis for submitting data that were incomplete or incorrect on the RED instrument (12 students and 10 hospital employees). Respondents ranged in age from 19 to 63, with a mean age of 30.74 years ($SD = 13.69$) and a median age of 23 years. Table 9 presents demographic characteristics for the final sample of 133.

Table 9
Summary of Study 2 Sample Characteristics

Sample Characteristics	Percentage ($N = 133$)
Gender	
Men	25.6
Women	72.2
No response	2.2
Educational level	
Less than high school	0.8
High school diploma or GED	50.4
2-year degree or technical school degree	16.5
4-year degree	18.8
Professional program degree	3.0
Master's degree or doctorate	10.5
College classification	
Not currently enrolled in college	44.3
Freshman	3.8
Sophomore	19.5
Junior	14.3
Senior	14.3
Master's student	1.7
No response	1.7

Materials. A revised version of RED measure was administered in Study 2. Instructions for the measure were revised to point out to respondents that their most likely reactions to each dilemma might not appear in RED's list of reaction statements. Participants were instructed to rank items based upon their relative significance to one another, even if their most likely reaction was not included in the list (see Appendix C for instructions).

In addition to the item rankings that were collected in the Study 1 version of RED, respondents were first asked to *rate* each statement on a 4-point scale. As in Study 1, the first set of items that appeared after each story were statements describing potential underlying issues, and these were rated as being either *very*, *somewhat*, *not very*, or *not at all significant*. The second and third sets of items for each story described potential interpersonal and intrapersonal reactions that people might experience in each dilemma, and respondents were asked to rate the likelihood that they would have such reactions (*very likely*, *somewhat likely*, *not very likely*, or *not at all likely*). After rating and ranking each of the response item sets for a story, participants were asked to select 4 of the 13 items that were most relevant or identifiable to them. Each of these four items were rank ordered as being *most identifiable*, *second most identifiable*, *third most identifiable*, or *fourth most identifiable*. Respondent instructions and a set of questions for one of the stories are shown in Appendix C.

To assess convergent validity, moral reasoning was measured with the DIT (Rest, 1975, 1979; Rest et al., 1999b). Similar to the structure of RED, the DIT presented respondents with a series of moral dilemmas, followed by issue statements to rate and rank. In the present study, a five-story version of the DIT was used. For each of five

dilemmas, participants rated 12 issue statements on the basis of their importance to the dilemma (*great, much, some, little, or no importance*). According to Rest et al. (1999b), each issue statement was designed to activate a moral schema representative of a particular stage (either Stage 2, 3, 4, 5a, 5b, or 6). The moral judgment stages represented by the DIT items were written to resemble Kohlberg's stages; the authors of the DIT indicated that there are some differences, and each item was designed to activate one of three schemas: the personal interests schema (Stages 2 and 3), the maintaining norms schema (Stage 4), and the postconventional schema (Stages 5a, 5b, and 6). Weighted DIT rankings provide scores for six stages of moral reasoning development, and a summary score indicating an individual's level of principled moral judgment (the P-Score). The P-Score is a weighted percentage that reflects the rankings participants give to the three highest stages of moral reasoning (Rest, 1990).

Dispositional optimism was measured with the LOT-R (Scheier, Carver, & Bridges, 1994). The LOT-R contained 10 statements that were rated using a five-point agreement scale (*0 = strongly disagree, 1 = disagree, 2 = neutral, 3 = agree, and 4 = strongly agree*). Of the 10 items, four were non-scored filler items. Scores for the six-item scale were summed, and scale scores could range from 0 to 24. Researchers have demonstrated that the LOT-R has acceptable internal consistency (.78). Studies have indicated that scores on the measure predict one's ability to cope and recover from physical illness and surgery (Scheier et al., 1989; Affleck et al., 2001), the psychological distress of coping with a disabled child (McLean, Harvey, Pallant, Bartlett, & Mutimer, 2004), and adaptation to college life (Brissette et al., 2002). The original version of the LOT contained eight scorable items. However, the authors removed two items (resulting

in the LOT-R) because, unlike the other six, they did not refer to an expectation of positive outcomes (Scheier et al., 1994).

Procedure. Student participants were recruited through flyers and announcements e-mailed to psychology professors offering extra credit for research participation. Students attended group data-collection sessions that lasted approximately 90 minutes. For hospital and public school employees, the researcher followed protocols that were negotiated with administrators from each organization. The researcher recruited public school employees by attending two school system events, a new teacher orientation and an employee benefits fair. At each event, the researcher manned a table with information about the study, questionnaire packets and gifts for participants. Employees who agreed to participate were given questionnaires and an information sheet in a self-addressed, postage-paid envelope, and they were instructed to complete the measures at home and return them to the researcher by mail. Employees were given a small gift at the time that they agreed to participate. Assessment packets were distributed to 60 school system employees and nine were returned, for a response rate of 15%.

For hospital employees, the researcher was given authorization to contact departmental managers to request permission to distribute assessments to their employees. At the discretion of department heads, a manager, shift supervisor or the researcher distributed questionnaire packets. Each envelope contained the measure, an information sheet explaining the study, and instructions for participating employees to complete the assessments at home and return them to a designated drop box. All respondents were assured that their participation was voluntary and that their responses would be anonymous. To protect the anonymity of hospital employees who participated,

the small gifts that were given for participation were left sitting out by the drop box for participants to take on the honor system. A total of 160 assessments were distributed, however it is not certain that managers and supervisors gave all of these to their employees. Therefore, the estimated response rate among hospital employees was *at least 41%*.

Results

Rating and Ranking Scales in the Revised Version of RED. As mentioned earlier, three types of scores were collected from the Study 2 version of RED. As in Study 1, participants ranked ordered three sets of items for each story, and these item rankings were reverse-coded so that high scores represented high item endorsement. Each item was also rated on a 4-point scale that ranged from *very significant* to *not at all significant*, or *very likely* to *not at all likely*. For the analysis, item ratings were coded from 1 (*not at all*) to 4 (*very*).

After rating and ranking all 13 items for a story, participants selected and rank-ordered four items that were most identifiable to them. From this final set of item rankings for each story, stage-level scores were calculated using a protocol that had been developed for scoring the DIT (Rest, 1990). For each story, Kegan stages were assigned points for items that were chosen among the top four. Stage scores were weighted in accordance with the ranking given to each item, with the highest ranked item receiving four points, three points for second place, two points for third, and one point for fourth. For example, imagine that a respondent selected a Stage 2 item as being the most identifiable statement in Story A, a Stage 3 item for second place, another Stage 2 item in third place, and a Stage 4 item in fourth place. The stage points accumulated in Story A

would be six points for Stage 2 (4 + 2), three points for Stage 3, and one point for Stage 4. To calculate a composite stage level score for all five stories combined, the points for each stage were summed and divided by .5, generating scores that could range from 0 to 100.

Omissions of Ratings and Rankings on RED. As mentioned above, 22 cases were removed from the analysis for submitting incomplete or incorrect forms. The criteria for removal from the study were not as rigorous in Study 2 as in Study 1 because the revised RED assessment contained three types of measurement scales. If respondents failed to complete the item rankings, but completed the item ratings, then their measures were included in the analysis of ratings (and vice-versa). Cases were removed on the basis of incomplete data if a participant skipped at least one set of both rankings and ratings. Several of the respondents who were excluded from the analysis provided the same rankings for multiple items throughout the measure. Of the 133 respondents who remained in the data set, participants who failed to complete sections of the test were tabulated to determine whether there were any patterns in the omissions, and these are presented in Table 10. Clearly, the sets of item rankings were most often omitted. Of the 15 sets of rankings (5 stories x 3 sets per story) the percentage of respondents who decided to skip a set of rankings ranged from 0.87% to 10.53%, and the intrapersonal sets (which contain five items, including a nonsensical item) were skipped more often than the issue statements and interpersonal reaction statements.

Table 10
Percentage of Participants Who Failed to Complete Specific Portions of the RED Measure in Study 2

Scale	Story A	Story B	Story C	Story D	Story E
Scales comprised of item rankings					
Issues	1.74%	3.5%	3.5%	2.61%	1.74%
Interpersonal Reactions	0.87%	1.74%	3.5%	2.61%	2.61%
Intrapersonal Reactions	5.22%	3.5%	6.96%	5.22%	10.43%
Scales comprised of item ratings					
Issues	--	--	--	0.87%	--
Interpersonal Reactions	--	--	1.74%	--	--
Intrapersonal Reactions	0.87%	--	1.74%	--	0.87%
Top 4 rankings for each story					
	0.87%	--	0.87%	1.74%	--

Descriptive Statistics and Comparison of RED Scores across Stories. Table 11 shows the average rankings and ratings for reaction statements within each story. Mean scores for each set of three stage-specific reaction statements within a story could range from 1 to 4.33. A repeated measures analysis of variance was conducted to determine whether there were any differences in the mean ratings from one story to the next. The purpose of this analysis was to determine whether each story and its response items could be considered parallel tests. Results of the ANOVA indicated that there were differences between stories for all of the scales, Stage 2: $F(4, 129) = 68.82$, Stage 3: $F(4, 129) =$

13.55, Stage 4: $F(4, 129) = 50.69$, Stage 5: $F(4, 129) = 38.86$, High Sounding: $F(4, 123) = 59.72$, all p -values $< .001$. Tables 12 through 21 show the distribution of ratings and rankings for RED's 65 individual items.

Table 11
Study 2 Mean and Median Stage Level Scores for Each Story from the RED Measure

Scale	Story A	Story B	Story C	Story D	Story E
Scales comprised of item rankings					
Stage 2: <i>M (SD)</i>	3.19 (.65)	2.68 (.65)	1.95 (.59)	2.51 (.70)	2.77 (.82)
<i>Mdn</i>	3.33	2.67	2.00	2.33	2.67
Stage 3: <i>M (SD)</i>	2.27 (.69)	2.80 (.58)	2.78 (.68)	3.15 (.67)	2.52 (.52)
<i>Mdn</i>	2.33	2.67	2.67	3.33	2.67
Stage 4: <i>M (SD)</i>	2.93 (.58)	3.36 (.64)	3.64 (.62)	2.87 (.58)	3.48 (.58)
<i>Mdn</i>	3.00	3.33	3.67	3.00	3.33
Stage 5: <i>M (SD)</i>	2.55 (.76)	2.20 (.76)	2.78 (.60)	2.46 (.70)	2.22 (.64)
<i>Mdn</i>	2.67	2.00	2.67	2.33	2.00
HS: <i>M (SD)</i>	2.06 (1.31)	1.87 (1.04)	1.35 (.75)	1.94 (1.24)	1.80 (1.19)
<i>Mdn</i>	2.00	2.00	1.00	1.00	1.00
Scales comprised of item ratings					
Stage 2: <i>M (SD)</i>	3.30 (.49)	2.93 (.58)	2.45 (.62)	2.54 (.59)	3.02 (.71)
<i>Mdn</i>	3.33	3.00	2.33	2.54	3.00
Stage 3: <i>M (SD)</i>	2.62 (.64)	2.95 (.64)	3.07 (.54)	2.98 (.73)	2.85 (.64)
<i>Mdn</i>	2.67	3.00	3.00	3.00	3.00
Stage 4: <i>M (SD)</i>	3.12 (.54)	3.38 (.52)	3.58 (.51)	2.92 (.49)	3.50 (.45)
<i>Mdn</i>	3.00	3.33	3.67	3.00	3.67
Stage 5: <i>M (SD)</i>	2.83 (.64)	2.58 (.64)	3.11 (.56)	2.39 (.63)	2.69 (.53)
<i>Mdn</i>	3.00	2.67	3.33	2.33	2.67
HS: <i>M (SD)</i>	2.32 (.91)	2.07 (.90)	1.50 (.70)	1.73 (.91)	2.38 (.94)
<i>Mdn</i>	2.00	2.00	1.00	1.00	2.00

Note. N ranged from 116 to 133. "HS" refers to high-sounding, nonsensical items.

Table 12
Study 2 Stage Level Rating Scores for Story A from the RED Measure (Frequencies and Percentages)

Items	Ratings				Total
	1 Not at all Significant	2 Not Very Significant	3 Somewhat Significant	4 Very Significant	
A: Issues Stage 2	3 2.3%	8 6.0%	23 17.3%	99 74.4%	133 100.0%
A: Issues Stage 3	7 5.3%	10 7.5%	59 44.4%	57 42.9%	133 100.0%
A: Issues Stage 4	11 8.3%	25 18.8%	54 40.6%	43 32.3%	133 100.0%
A: Issues Stage 5	10 7.5%	23 17.3%	65 48.9%	35 26.3%	133 100.0%
A: Interpersonal Stage 2	6 4.5%	9 6.8%	52 39.1%	66 49.6%	133 100.0%
A: Interpersonal Stage 3	50 37.6%	40 30.1%	28 21.1%	15 11.3%	133 100.0%
A: Interpersonal Stage 4	5 3.8%	22 16.5%	47 35.3%	59 44.4%	133 100.0%
A: Interpersonal Stage 5	18 13.5%	34 25.6%	55 41.4%	26 19.5%	133 100.0%
A: Intrapersonal Stage 2	15 11.4%	22 16.7%	54 40.9%	41 31.1%	132 100.0%
A: Intrapersonal Stage 3	22 16.7%	37 28.0%	52 39.4%	21 15.9%	132 100.0%
A: Intrapersonal Stage 4	5 3.8%	13 9.8%	68 51.5%	46 34.8%	132 100.0%
A: Intrapersonal Stage 5	15 11.4%	34 25.8%	35 26.5%	48 36.4%	132 100.0%
A: Intrapersonal High Sounding	27 20.5%	49 37.1%	43 32.6%	13 9.8%	132 100.0%

Table 13
Study 2 Stage Level Rating Scores for Story B from the RED Measure (Frequencies and Percentages)

Items	Ratings				Total
	1 Not at all Significant	2 Not Very Significant	3 Somewhat Significant	4 Very Significant	
B: Issues Stage 2	3 2.3%	5 3.8%	32 24.1%	93 69.9%	133 100.0%
B: Issues Stage 3	7 5.3%	25 18.8%	58 43.6%	43 32.3%	133 100.0%
B: Issues Stage 4	10 7.5%	27 20.3%	52 39.1%	44 33.1%	133 100.0%
B: Issues Stage 5	16 12.0%	45 33.8%	53 39.8%	19 14.3%	133 100.0%
B: Interpersonal Stage 2	25 18.8%	42 31.6%	46 34.6%	20 15.0%	133 100.0%
B: Interpersonal Stage 3	26 19.5%	31 23.3%	51 38.3%	25 18.8%	133 100.0%
B: Interpersonal Stage 4	1 .8%	9 6.8%	24 18.0%	99 74.4%	133 100.0%
B: Interpersonal Stage 5	25 18.8%	30 22.6%	49 36.8%	29 21.8%	133 100.0%
B: Intrapersonal Stage 2	17 12.8%	39 29.3%	43 32.3%	34 25.6%	133 100.0%
B: Intrapersonal Stage 3	8 6.0%	16 12.0%	42 31.6%	67 50.4%	133 100.0%
B: Intrapersonal Stage 4	6 4.5%	4 3.0%	41 30.8%	82 61.7%	133 100.0%
B: Intrapersonal Stage 5	27 20.3%	31 23.3%	49 36.8%	26 19.5%	133 100.0%
B: Intrapersonal High Sounding	45 34.1%	37 28.0%	46 34.8%	4 3.0%	132 100.0%

Table 14
Study 2 Stage Level Rating Scores for Story C from the RED Measure (Frequencies and Percentages)

Items	Ratings				Total
	1 Not at all Significant	2 Not Very Significant	3 Somewhat Significant	4 Very Significant	
C: Issues Stage 2	11 8.3%	27 20.3%	42 31.6%	53 39.8%	133 100.0%
C: Issues Stage 3	2 1.5%	12 9.0%	41 30.8%	78 58.6%	133 100.0%
C: Issues Stage 4	1 .8%	11 8.3%	24 18.0%	97 72.9%	133 100.0%
C: Issues Stage 5	4 3.0%	12 9.0%	55 41.4%	62 46.6%	133 100.0%
C: Interpersonal Stage 2	27 20.6%	44 33.6%	42 32.1%	18 13.7%	131 100.0%
C: Interpersonal Stage 3	26 19.8%	51 38.9%	40 30.5%	14 10.7%	131 100.0%
C: Interpersonal Stage 4	4 3.1%	1 .8%	34 26.0%	92 70.2%	131 100.0%
C: Interpersonal Stage 5	12 9.2%	23 17.6%	59 45.0%	37 28.2%	131 100.0%
C: Intrapersonal Stage 2	47 35.9%	57 43.5%	19 14.5%	8 6.1%	131 100.0%
C: Intrapersonal Stage 3	2 1.5%	8 6.1%	55 42.0%	66 50.4%	131 100.0%
C: Intrapersonal Stage 4	4 3.1%	8 6.1%	41 31.3%	78 59.5%	131 100.0%
C: Intrapersonal Stage 5	2 1.5%	21 16.0%	69 52.7%	39 29.8%	131 100.0%
C: Intrapersonal High Sounding	78 59.5%	42 32.1%	9 6.9%	2 1.5%	131 100.0%

Table 15
Study 2 Stage Level Rating Scores for Story D from the RED Measure (Frequencies and Percentages)

Items	Ratings				Total
	1 Not at all Significant	2 Not Very Significant	3 Somewhat Significant	4 Very Significant	
D: Issues Stage 2	53 40.2%	36 27.3%	27 20.5%	16 12.1%	132 100.0%
D: Issues Stage 3	15 11.4%	24 18.2%	52 39.4%	41 31.1%	132 100.0%
D: Issues Stage 4	2 1.5%	8 6.1%	39 29.5%	83 62.9%	132 100.0%
D: Issues Stage 5	41 31.1%	56 42.4%	28 21.2%	7 5.3%	132 100.0%
D: Interpersonal Stage 2	14 10.5%	31 23.3%	69 51.9%	19 14.3%	133 100.0%
D: Interpersonal Stage 3	19 14.3%	30 22.6%	49 36.8%	35 26.3%	133 100.0%
D: Interpersonal Stage 4	6 4.5%	15 11.3%	44 33.1%	68 51.1%	133 100.0%
D: Interpersonal Stage 5	9 6.8%	20 15.0%	47 35.3%	57 42.9%	133 100.0%
D: Intrapersonal Stage 2	16 12.0%	25 18.8%	54 40.6%	38 28.6%	133 100.0%
D: Intrapersonal Stage 3	6 4.5%	11 8.3%	55 41.4%	61 45.9%	133 100.0%
D: Intrapersonal Stage 4	47 35.3%	53 39.8%	29 21.8%	4 3.0%	133 100.0%
D: Intrapersonal Stage 5	50 37.6%	39 29.3%	37 27.8%	7 5.3%	133 100.0%
D: Intrapersonal High Sounding	72 54.1%	31 23.3%	24 18.0%	6 4.5%	133 100.0%

Table 16
Study 2 Stage Level Rating Scores for Story E from the RED Measure (Frequencies and Percentages)

Items	Ratings				Total
	1 Not at all Significant	2 Not Very Significant	3 Somewhat Significant	4 Very Significant	
E: Issues Stage 2	15 11.3%	23 17.3%	51 38.3%	44 33.1%	133 100.0%
E: Issues Stage 3	19 14.3%	32 24.1%	52 39.1%	30 22.6%	133 100.0%
E: Issues Stage 4	2 1.5%	1 .8%	32 24.1%	98 73.7%	133 100.0%
E: Issues Stage 5	8 6.0%	27 20.3%	69 51.9%	29 21.8%	133 100.0%
E: Interpersonal Stage 2	21 15.8%	28 21.1%	45 33.8%	39 29.3%	133 100.0%
E: Interpersonal Stage 3	22 16.5%	30 22.6%	45 33.8%	36 27.1%	133 100.0%
E: Interpersonal Stage 4	7 5.3%	15 11.3%	43 32.3%	68 51.1%	133 100.0%
E: Interpersonal Stage 5	17 12.8%	46 34.6%	59 44.4%	11 8.3%	133 100.0%
E: Intrapersonal Stage 2	4 3.0%	16 12.1%	39 29.5%	73 55.3%	132 100.0%
E: Intrapersonal Stage 3	7 5.3%	16 12.2%	59 45.0%	49 37.4%	131 100.0%
E: Intrapersonal Stage 4	2 1.5%	12 9.1%	35 26.5%	83 62.9%	132 100.0%
E: Intrapersonal Stage 5	12 9.2%	36 27.5%	64 48.9%	19 14.5%	131 100.0%
E: Intrapersonal High Sounding	26 20.0%	44 33.8%	44 33.8%	16 12.3%	130 100.0%

Table 17
Study 2 Stage Level Ranking Scores for Story A from the RED Measure (Frequencies and Percentages)

Items	Rankings (Reverse Coded)					Total
	Least Preferred				Most Preferred	
	1	2	3	4	5	
A: Issues Stage 2	18 13.7%	12 9.2%	17 13.0%	84 64.1%	-- --	131 100.0%
A: Issues Stage 3	24 18.3%	39 29.8%	48 36.6%	20 15.3%	-- --	131 100.0%
A: Issues Stage 4	35 26.7%	45 34.4%	38 29.0%	13 9.9%	-- --	131 100.0%
A: Issues Stage 5	54 41.2%	35 26.7%	28 21.4%	14 10.7%	-- --	131 100.0%
A: Interpersonal Stage 2	11 8.3%	16 12.1%	54 40.9%	51 38.6%	-- --	132 100.0%
A: Interpersonal Stage 3	87 65.9%	24 18.2%	11 8.3%	10 7.6%	-- --	132 100.0%
A: Interpersonal Stage 4	15 11.4%	38 28.8%	28 21.2%	51 38.6%	-- --	132 100.0%
A: Interpersonal Stage 5	19 14.4%	54 40.9%	39 29.5%	20 15.2%	-- --	132 100.0%
A: Intrapersonal Stage 2	14 11.1%	27 21.4%	29 23.0%	28 22.2%	28 22.2%	126 100.0%
A: Intrapersonal Stage 3	21 16.7%	34 27.0%	35 27.8%	21 16.7%	15 11.9%	126 100.0%
A: Intrapersonal Stage 4	4 3.2%	19 15.1%	27 21.4%	33 26.2%	43 34.1%	126 100.0%
A: Intrapersonal Stage 5	25 19.8%	20 15.9%	19 15.1%	32 25.4%	30 23.8%	126 100.0%
A: Intrapersonal High Sounding	62 49.2%	26 20.6%	16 12.7%	12 9.5%	10 7.9%	126 100.0%

Note. Only the intrapersonal item set included 5 items.

Table 18

Study 2 Stage Level Ranking Scores for Story B from the RED Measure (Frequencies and Percentages)

Items	Rankings (Reverse Coded)					Total
	Least Preferred				Most Preferred	
	1	2	3	4	5	
B: Issues Stage 2	7 5.5%	15 11.7%	37 28.9%	69 53.9%	--	128 100.0%
B: Issues Stage 3	26 20.3%	42 32.8%	41 32.0%	19 14.8%	--	128 100.0%
B: Issues Stage 4	34 26.8%	33 26.0%	37 29.1%	23 18.1%	--	127 100.0%
B: Issues Stage 5	62 48.4%	37 28.9%	12 9.4%	17 13.3%	--	128 100.0%
B: Interpersonal Stage 2	33 25.2%	52 39.7%	39 29.8%	7 5.3%	--	131 100.0%
B: Interpersonal Stage 3	36 27.5%	46 35.1%	40 30.5%	9 6.9%	--	131 100.0%
B: Interpersonal Stage 4	10 7.6%	5 3.8%	20 15.3%	96 73.3%	--	131 100.0%
B: Interpersonal Stage 5	52 39.7%	28 21.4%	32 24.4%	19 14.5%	--	131 100.0%
B: Intrapersonal Stage 2	31 24.2%	27 21.1%	42 32.8%	20 15.6%	8 6.3%	128 100.0%
B: Intrapersonal Stage 3	5 3.9%	10 7.8%	27 21.1%	44 34.4%	42 32.8%	128 100.0%
B: Intrapersonal Stage 4	3 2.3%	9 7.0%	16 12.5%	36 28.1%	64 50.0%	128 100.0%
B: Intrapersonal Stage 5	28 22.0%	41 32.3%	27 21.3%	19 15.0%	12 9.4%	127 100.0%
B: Intrapersonal High Sounding	60 46.9%	40 31.3%	16 12.5%	9 7.0%	3 2.3%	128 100.0%

Note. Only the intrapersonal item set included 5 items.

Table 19

Study 2 Stage Level Ranking Scores for Story C from the RED Measure (Frequencies and Percentages)

Items	Rankings (Reverse Coded)					Total
	Least Preferred				Most Preferred	
	1	2	3	4	5	
C: Issues Stage 2	62 48.1%	27 20.9%	32 24.8%	8 6.2%	--	129 100.0%
C: Issues Stage 3	25 19.4%	41 31.8%	28 21.7%	35 27.1%	--	129 100.0%
C: Issues Stage 4	7 5.4%	23 17.8%	26 20.2%	73 56.6%	--	129 100.0%
C: Issues Stage 5	35 27.1%	38 29.5%	43 33.3%	13 10.1%	--	129 100.0%
C: Interpersonal Stage 2	60 46.5%	40 31.0%	18 14.0%	11 8.5%	--	129 100.0%
C: Interpersonal Stage 3	45 34.9%	48 37.2%	32 24.8%	4 3.1%	--	129 100.0%
C: Interpersonal Stage 4	3 2.3%	12 9.3%	21 16.3%	93 72.1%	--	129 100.0%
C: Interpersonal Stage 5	21 16.3%	29 22.5%	58 45.0%	21 16.3%	--	129 100.0%
C: Intrapersonal Stage 2	26 21.0%	71 57.3%	17 13.7%	7 5.6%	3 2.4%	124 100.0%
C: Intrapersonal Stage 3	1 .8%	9 7.3%	38 30.6%	26 21.0%	50 40.3%	124 100.0%
C: Intrapersonal Stage 4	4 3.2%	10 8.1%	12 9.7%	46 37.1%	52 41.9%	124 100.0%
C: Intrapersonal Stage 5	1 .8%	8 6.5%	55 44.4%	43 34.7%	17 13.7%	124 100.0%
C: Intrapersonal High Sounding	92 74.2%	26 21.0%	2 1.6%	2 1.6%	2 1.6%	124 100.0%

Note. Only the intrapersonal item set included 5 items.

Table 20

Study 2 Stage Level Ranking Scores for Story D from the RED Measure (Frequencies and Percentages)

Items	Rankings (Reverse Coded)					Total
	Least Preferred				Most Preferred	
	1	2	3	4	5	
D: Issues Stage 2	65	37	20	8	--	130
	50.0%	28.5%	15.4%	6.2%	--	100.0%
D: Issues Stage 3	10	24	63	33	--	130
	7.7%	18.5%	48.5%	25.4%	--	100.0%
D: Issues Stage 4	3	17	26	84	--	130
	2.3%	13.1%	20.0%	64.6%	--	100.0%
D: Issues Stage 5	52	52	21	5	--	130
	40.0%	40.0%	16.2%	3.8%	--	100.0%
D: Interpersonal Stage 2	42	35	33	20	--	130
	32.3%	26.9%	25.4%	15.4%	--	100.0%
D: Interpersonal Stage 3	41	32	37	20	--	130
	31.5%	24.6%	28.5%	15.4%	--	100.0%
D: Interpersonal Stage 4	20	31	38	41	--	130
	15.4%	23.8%	29.2%	31.5%	--	100.0%
D: Interpersonal Stage 5	27	32	22	49	--	130
	20.8%	24.6%	16.9%	37.7%	--	100.0%
D: Intrapersonal Stage 2	13	13	21	49	30	126
	10.3%	10.3%	16.7%	38.9%	23.8%	100.0%
D: Intrapersonal Stage 3	5	7	10	27	77	126
	4.0%	5.6%	7.9%	21.4%	61.1%	100.0%
D: Intrapersonal Stage 4	24	52	32	15	3	126
	19.0%	41.3%	25.4%	11.9%	2.4%	100.0%
D: Intrapersonal Stage 5	15	34	45	21	11	126
	11.9%	27.0%	35.7%	16.7%	8.7%	100.0%
D: Intrapersonal High Sounding	70	19	17	15	5	126
	55.6%	15.1%	13.5%	11.9%	4.0%	100.0%

Note. Only the intrapersonal item set included 5 items.

Table 21

Study 2 Stage Level Ranking Scores for Story E from the RED Measure (Frequencies and Percentages)

Items	Rankings (Reverse Coded)					Total
	Least Preferred				Most Preferred	
	1	2	3	4	5	
E: Issues Stage 2	40 30.8%	37 28.5%	28 21.5%	25 19.2%	--	130 100.0%
E: Issues Stage 3	52	38 29.2%	31 23.8%	9 6.9%	--	130 100.0%
E: Issues Stage 4	4	11 8.5%	26 20.0%	89 68.5%	--	130 100.0%
E: Issues Stage 5	34	44 33.8%	45 34.6%	7 5.4%	--	130 100.0%
E: Interpersonal Stage 2	29	34 26.2%	37 28.5%	30 23.1%	--	130 100.0%
E: Interpersonal Stage 3	32	44 33.8%	29 22.3%	25 19.2%	--	130 100.0%
E: Interpersonal Stage 4	12	23 17.7%	32 24.6%	63 48.5%	--	130 100.0%
E: Interpersonal Stage 5	57	29 22.3%	32 24.6%	12 9.2%	--	130 100.0%
E: Intrapersonal Stage 2	12	17 14.3%	23 19.3%	25 21.0%	42 35.3%	119 100.0%
E: Intrapersonal Stage 3	12	18 15.0%	27 22.5%	42 35.0%	21 17.5%	120 100.0%
E: Intrapersonal Stage 4	4	13 10.8%	27 22.5%	31 25.8%	45 37.5%	120 100.0%
E: Intrapersonal Stage 5	21	47 39.2%	32 26.7%	15 12.5%	5 4.2%	120 100.0%
E: Intrapersonal High Sounding	70 58.3%	25 20.8%	11 9.2%	7 5.8%	7 5.8%	120 100.0%

Note. Only the intrapersonal item set included 5 items.

Reliability Analysis of RED. Cronbach's alpha reliability scores are shown in Table 22 for stage-level scales comprised of item rankings, item ratings, and for the weighted percentage scores generated from each story's top four rankings. As in Study 1, alpha coefficients for item rankings were optimal when respondents were screened out if they highly rated or ranked at least two high-sounding items. However, this filter improved reliability only slightly for some scales, while slightly reducing it for others.

Table 22
Study 2 Comparisons of Cronbach Alpha Reliabilities of Stage Scales Before and After Removing Cases and Items for the RED Measure

Scale	Total Sample, All Items	Sample Endorsing < 2 High Sounding Items, All Items	Sample Endorsing < 2 High Sounding Items, Scales with Low Performing Items Removed
Scales comprised of item rankings			
Stage 2	.53	.55	.61
Stage 3	.44	.46	.54
Stage 4	.37	.36	.43
Stage 5	.63	.65	.66
Scales comprised of item ratings			
Stage 2	.73	.74	--
Stage 3	.74	.74	--
Stage 4	.63	.64	--
Stage 5	.70	.69	--
Scales comprised of weighted percentage scores from top 4 rankings			
Stage 2	.47	.48	---
Stage 3	.22	.24	---
Stage 4	.35	.34	---
Stage 5	.53	.53	---

Note. *N*'s ranged from 99 to 103 for item rankings, from 122 to 128 for item ratings, and from 125 to 131 for top 4 rankings.

Of the three measurement scales presented in Table 22, three of the four rating scales reached the acceptable alpha level of .70 or greater (Nunnally, 1978), ranging from .64 to .74. Alpha levels for the ranking scales fell below the minimum acceptable level, particularly for the scales comprised of weighted percentage scores from top four rankings. Each of these stage-level scales consisted of only five items, one per story, and alpha levels ranged from .24 to .53 after conducting the filter for nonsensical items. For the 15-item ranking scales, reliability coefficients ranged from .43 to .66 after the filter for nonsensical items and removing two items each from scales for Stages 2 through 4, and one item from the Stage 5 scale.

DIT Scores. The scoring protocol outlined in the DIT manual (Rest, 1979) was used to generate stage scores and the P-Score for the DIT. The P-Score is the most common index for the DIT, and represents the extent to which individuals endorse higher levels of moral judgment. Only the DITs three highest stage scales contribute to this score (stages 5a, 5b, and 6).

Rest (1990) recommended three “reliability checks” to identify DITs that should be removed from analyses. The first reliability check was to assess the number of ranking points assigned to the measure’s nonsensical items (referred to as the “M-Score”). If a participant assigned more than 14% of the measure’s ranking points to nonsensical items, the subject was removed from the analysis. The second consistency check was a comparison of rankings and ratings for each story. In the DIT instructions, participants were told to rate each of twelve issues on a four-point scale, and then select their top four issue statements and rank them. Items included in the top-four rankings for each story should be the highest rated items. If the items that a participant selected for

first and second place were not the most highly rated, then this was considered to be an inconsistency between ratings and rankings. Participants were removed from the analysis if their ratings and rankings were inconsistent on at least two stories. From these two consistency checks combined, 26 respondents were removed from the DIT analysis in addition to the participants who had already been removed for providing incomplete data in the RED measure.

The final consistency check was to identify respondents who did not differentiate sufficiently between items in their ratings (e.g., at least 10 out of 12 statements given the same rating on two or more dilemmas). No participants needed to be removed on the basis of this screening procedure. According to Rest (1990), it is normal to lose between 5% and 15% of respondents as a result of these consistency checks. In the present study, the percentage of respondents removed from the analysis was slightly higher (16.8%).

To calculate internal consistency for the DIT, stage-level scores and P-Scores were calculated for each of the five dilemmas, yielding a five-item scale for each stage level and the P-Score. Cronbach alpha coefficients were low. For the P-Score, the reliability coefficient was .54, and individual scale internal consistency estimates were .33 (Stage 2), .42 (Stage 3), .50 (Stage 4), .45 (Stage 5a), .31 (Stage 5b), and .17 (Stage 6). According to Rest (1979), average internal consistency for the P-Score is .77, and coefficients for the individual stage scales range from .28 to .60. Therefore, the reliability coefficients in the present study (.17 to .50) were lower than usual.

Table 23 presents mean scores and standard deviations from the present sample compared with normative data from Rest (1990). To compare the present sample with the normative data, the sample was divided into two groups, college students and college

graduates. In the present study, the pattern of mean scores across the DIT's six stage scales was similar to the reported norms. P-Scores were approximately seven points lower in the present study than the reported norms for each subgroup.

Table 23
Comparison of Study 2 Defining Issues Test (DIT) Scores with Normative Data

	DIT Scores						
	2	3	4	5a	5b	6	P
College Students							
Norm	3.05 (2.81)	8.60 (5.14)	17.01 (8.07)	15.81 (6.31)	5.20 (3.40)	4.89 (3.34)	43.19 (14.32)
Present Study	2.84 (2.85)	8.07 (5.25)	16.71 (6.78)	13.55 (6.20)	2.29 (2.12)	2.53 (2.34)	36.48 (15.86)
Graduates*							
Norm	2.24 (2.51)	7.96 (5.66)	17.97 (8.67)	15.09 (6.11)	5.26 (3.52)	6.56 (3.35)	44.85 (15.06)
Present Study	1.94 (2.27)	5.75 (4.00)	19.42 (7.31)	11.00 (5.01)	2.65 (2.86)	4.56 (2.62)	37.15 (13.90)

Note. Standard deviations are shown in parentheses. The P-Score represents a weighted percentage score for Stages 5a, 5b, and 6.

* As reported in Rest (1990), the normative sample of graduates was comprised of college graduates who were not in graduate school. Data from the present sample of hospital or public school employees included graduate students.

Correlation Analyses. To assess scale intercorrelations and construct validity, a Spearman rho correlation analysis was performed for scores on RED, the DIT, age, education level, and college classification among. To assess discriminant validity, a correlation analysis was also conducted to assess the relationship between scores on RED and the LOT-R. Table 24 presents correlation coefficients for these variables and RED's ranking and rating scales. Correlations between the construct validity variables and

RED's item ranking scales are shown below the table diagonal. The coefficients above the diagonal are for correlations using the RED rating scales.

One of the differences between rating and ranking scales is that rated items and subscales are likely to correlate positively with one another and with other variables, while negative correlations are essentially built-in when using ranks (Alwin & Krosnic, 1985; Krosnick & Alwin, 1988). For scale intercorrelations, negative coefficients are inherent because a high ranking on one item results in a lower ranking on another. This difference was evident from the results reported in Table 24. Most of the RED stage-level rating scales were positively intercorrelated at $p < .05$ or better, and the relative strength of correlations generally fit the theoretical distance between stages. However, correlations between adjacent stages at each end of the continuum were strongest. For example, Stages 2 and 3 had a high correlation of .688 ($p < .01$) and the correlation between Stages 4 and 5 was .535 ($p < .01$).

Intercorrelations among the ranked scales were similar to those observed in Study 1. Most of the intercorrelations were negative, and the one significant positive correlation was between the adjacent Stages 2 and 3 ($r_s = .195, p < .05$).

There were a few significant correlations between the DIT and RED rating scale scores that supported RED's construct validity. However, correlations between the DIT and scores from the RED ranking scales revealed a stronger pattern of relationships. The DIT Stage 3 score correlated significantly with the RED scores for Stage 2 ($r_s = .322, p < .01$), Stage 4 ($r_s = -.231, p < .05$), and Stage 5 ($r_s = -.265, p < .01$). The DIT Stage 6 score correlated significantly with all of the RED ranking scales, revealing a negative relationship with the lower scales and a positive relationship with higher scales (Stage 2:

$r_s = -.211, p < .05$, Stage 3: $r_s = -.212, p < .05$, Stage 4: $r_s = .293, p < .01$; Stage 5: $r_s = .248, p < .05$). Two correlations between the P-score and RED were significant, and these were the RED ranking scores for Stage 4 ($r_s = .229, p < .05$) and Stage 2 ($r_s = -.235, p < .05$).

The pattern of scores between RED and age, educational level, and college classification followed a pattern similar to the relationships observed in Study 1. Overall, the ranking scales yielded a higher number of significant correlations than the ratings.

A Spearman rho correlation analysis comparing DIT scores to the weighted percentage scores from RED's top four rankings revealed a similar, yet weaker pattern of significant correlations. During the analysis, a typographical error was discovered in the section of the questionnaire in which respondents were to rank order their top four reaction statements. Each reaction statement for a dilemma was labeled from "a" to "m" in lower-case letters typed in Arial font. To select their top four reaction statements for each story, participants were presented with four rows of letters, and circled the letter corresponding to their first, second, third, and fourth choices (see Appendix C). However, the word processor's autocorrect function capitalized the letter "i," making it identical to the lower-case "l." As a result, respondents who intended to circle the letter "l" may have mistakenly selected the capital "i." Therefore, data collected from the top four rankings probably contains errors and results should be viewed with caution.

Table 24

Study 2 Bivariate Correlations among Rating and Ranking Scales for the RED Measure, DIT, Age, Education, College Class, and LOT-R

	RED 2	RED 3	RED 4	RED 5	DIT 2	DIT 3	DIT 4	DIT 5a	DIT 5b	DIT 6	P-Score	Age	Ed	Class	LOT
RED 2	--	.688**	.285**	.168	.029	.161	.160	-.070	-.162	-.295**	-.184	-.329**	-.259**	-.191	-.307**
RED 3	.195*	--	.280**	.216*	.086	.068	.192*	-.052	-.103	-.293**	-.179	-.317**	-.183*	-.314**	-.235**
RED 4	-.391**	-.388**	--	.535**	-.144	-.114	.088	.030	.051	-.092	-.007	.104	.168	-.135	-.024
RED 5	-.649**	-.578**	.166	--	-.116	-.179	-.057	.178	.200*	-.079	.123	.007	.105	-.106	-.036
DIT 2	.073	.115	-.173	-.043	--	.087	-.012	-.313**	-.203*	-.168	-.385**	-.320**	-.275**	-.331**	-.083
DIT 3	.322**	.094	-.231*	-.265**		--	-.443**	-.045	-.034	-.397**	-.160	-.276**	-.173	-.135	-.097
DIT 4	.048	.156	-.020	-.054			--	-.606**	-.338**	-.032	-.612**	.163	.158	-.094	.039
DIT 5a	-.188	-.097	.094	.111				--	.200*	-.039	.848**	-.093	-.067	.211	-.076
DIT 5b	-.170	-.033	.131	.101					--	.168	.472**	.095	.267**	-.009	.248*
DIT 6	-.211*	-.212*	.293**	.248*						--	.390**	.414**	.285**	.104	.225*
DIT P	-.235*	-.148	.229*	.161							--	.128	.112	.217	.101
Age	-.346**	-.292**	.327**	.327**								--	.720**	.834**	.252**
Ed	-.267**	-.199*	.220*	.250**									--	.454**	.204*
Class	-.076	-.295*	.066	.212										--	.031
LOT-R	-.264**	-.025	.204*	.111											--

Note. Correlation coefficients were calculated using the Spearman rho procedure for ordinal data. Coefficients above the diagonal were calculated with RED's nonipsative rating scales. Coefficients below the diagonal were calculated with RED's ipsative ranking scales. Sample sizes ranged from 89 to 111 for all pairs except for correlations with college classification, where the sample of college students ranged from 54 to 63.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Discriminant Validity. It was hypothesized that scores from RED would be uncorrelated with scores from the LOT-R. As mentioned earlier, the LOT-R is a measure of dispositional optimism and no literature was found to indicate that this construct would be related to an individual's level of meaning-making. From the 10 items included on the scale, four filler items were removed and two items were reverse-coded so that high ratings on each item represented high optimism. Scores on the LOT-R were summed and could range from 0 to 24. A Cronbach's alpha reliability analysis was performed and the alpha level was .79.

Contrary to the hypothesis, scores from the LOT-R correlated negatively with RED Stage 2 ($r_s = -.307, p < .01$) and Stage 3 ($r_s = -.235, p < .01$) from the rating scales. From the RED ranking scales, the LOT-R correlated negatively with Stage 2 ($r_s = -.264, p < .01$) and positively with Stage 4 ($r_s = .204, p < .05$). However, there were also unexpected positive correlations between the LOT-R with age ($r_s = .252, p < .01$) and education ($r_s = .204, p < .05$). Given the LOT-R's correlation with age and education, it did not appear to have been an appropriate choice for establishing discriminant validity in this study because these two variables were not expected to be correlated, and age is correlated with developmental level for several DIT and RED scales.

Because RED was based upon a developmental construct, it was expected that scores would correlate with age, educational level, and the DIT, another developmental construct. However, the measure's correlation with age and education could also raise a question about its validity. Did RED correlate with the DIT and the LOT-R because it truly measured a construct related to each of these assessments, or were age and educational level driving the correlations?

To test for the effects of age and education on RED's relationship with the DIT and the LOT-R, a nonparametric partial correlation analysis was conducted that controlled for the potential effects of these variables. Table 25 shows correlation coefficients between the RED ranking scales without this control (in left-hand columns) compared to the partial correlations (in the right-hand columns). The result was a loss of 6 out of 11 significant relationships, and a weakening of the remaining 5 significant correlations between the scales scores from RED and the DIT. Two previously non-significant correlation coefficients with the DIT 5 scale became significant, a negative correlation between the DIT 5a and RED Stage 2 scales ($r_s = -.230, p < .05$), and a positive correlation between the DIT 4 and RED Stage 3 scales ($r_s = .215, p < .05$). As for the relationship between RED and the LOT-R, one previously significant correlation disappeared (for the RED Stage 4 scale), and the correlation between Red Stage 2 and the LOT-R weakened.

Table 25
Correlation Coefficients between the RED Ranking Scales, DIT, and LOT Before and After Controlling for Age and Education

Scales	Correlation Coefficients Before Controlling for Age and Education				Partial Correlation Coefficients After Controlling for Age and Education			
	RED 2	RED 3	RED 4	RED 5	RED 2	RED 3	RED 4	RED 5
DIT 2	.073	.115	-.173	-.043	-.002	.106	-.117	-.000
DIT 3	.322**	.094	-.231*	-.265**	.273*	.008	-.184	-.203*
DIT 4	.048	.156	-.020	-.054	.086	.215*	-.052	-.094
DIT 5a	-.188	-.097	.094	.111	-.230*	-.142	.117	.154
DIT 5b	-.170	-.033	.131	.101	-.141	-.034	.098	.108
DIT 6	-.211*	-.212*	.293**	.248*	-.096	-.084	.198*	.122
DIT P	-.235*	-.148	.229*	.161	-.205*	-.124	.189	.131
LOT	-.264**	-.025	.204*	.111	-.213*	.111	.078	.039

Study 2 Discussion

Study 2 moderately supported Hypothesis 1. Stage level scores from RED correlated with several stage scores of the DIT and the P-Score in a pattern compatible with the stage structures of each measure. Positive correlations between RED's ranking scales and the DIT typically fit the theoretical correspondence between Kohlberg's stages of moral judgment and Kegan's stages of meaning-making. For example, after controlling for age and education, the RED 2 ranking score was positively related to the DIT 3 score, and RED 4 correlated positively with DIT 5a. The RED Stage 4 score also correlated positively with the P-Score. However, it should be noted that the DIT Manual (Rest, 1990) strongly discourages stage typing from the DIT. Instead, each stage-level serves as an indication of an individual's tendency to endorse the type of reasoning associated with each stage. In addition to correlations with the DIT, RED scales correlated significantly with age and education in a pattern that made theoretical sense, particularly among the ranking scales.

Study 2 results did not completely support Hypothesis 2. The LOT-R correlated negatively with RED Stage 2, even after controlling for age and education level. However, partial correlations between the LOT-R and Stages 3 through 5 were non-significant.

While reliability scores for the rating scales were fairly high, the ranking scale alpha coefficients remained low. However, despite the low alpha coefficients, the ranking scales produced the strongest correlations between RED and the construct validity variables. The reliability coefficients for the DIT were also lower than estimates that have been published in the past.

IV. GENERAL DISCUSSION

The present study was a fairly successful first step toward the development of a quantitative measure of how people construct meaning. RED's ranking scales, and to a lesser extent its rating scales, correlate with several scales from a measure of moral reasoning, a construct that has been empirically linked to subject-object interview scores (Pratt et al., 1991). In addition, the measure's stage-level scales correlate with each other, age and education in a pattern that fits Kegan's proposed stage sequence. In Pratt et al.'s (1991) comparison of subject-object scores with the weighted average scores from Kohlberg's moral judgment interview, the correlation between these two scores was .42 without controlling for age or education. However, both scores were derived from the same interview data, and the researchers applied different scoring procedures to obtain the subject-object score and the weighted average score for moral reasoning. Given the common method variance, this value can be considered the upper limit of the relationship between these two developmental constructs. In the present study, more modest correlation coefficients were obtained in the .20s and .30s. However, these relationships were achieved by comparing separate measures for each construct. Therefore, after undergoing several revisions that are described below, research with RED should be continued to solidly establish construct validity, as well as criterion-related validity.

In the present study, low internal consistency estimates were obtained for both RED rankings and the DIT. Sample characteristics were probably, in part, responsible

for the low reliability levels. In both studies, most participants were in their 20's or younger. While the Study 2 sample was somewhat more diverse in age than Study 1, the age distribution remained heavily skewed toward younger ages. In addition to the high percentage of young participants in Study 2 (60.9% under the age of 31), 17.3% were older than 50. Therefore, only fifth of the sample (21.8%) fell within the 20-year age range of 31 to 50 years of age. During this period of adult life, many individuals are operating from Kegan's Stage 3, or are making the Stage 3 to Stage 4 transition (Kegan, 1994), and these individuals were probably under-represented in the sample. Future research should be conducted with a sample that provides better age group representation.

The next step toward developing RED into a valid measure of meaning-making should be an analysis of the open-ended reaction statements that participants wrote for each story. These narratives may provide ideas for the development of new items and revisions for current items.

There was a potentially confounding factor that could not be ignored as an explanation for the relationship between RED and the other variables of interest. The potential confound was the reading level of the response items. When developing these items, an attempt was made to write all statements at a similar reading level. However, it soon became obvious that increases in the complexity of meaning-making as individuals move from stage to stage are difficult to describe without increasing the complexity of the writing style. The higher-level stage items were not always written in a higher reading level, but variability in the difficulty of items throughout the measure could have had an impact upon responses. In a review of the reading levels of normal personality measures, Schinka and Borum (1994) reported that personality measures typically

required a reading level between the sixth and eighth grades, and the authors expressed concern that items written at the 8th grade level might pose a problem for some respondents. At present, reading levels of the statements on RED range from 5th to 12th grade.

In Study 2, an examination of item omissions revealed that participants were more likely to skip the intrapersonal rankings than any other portion of the test. Although the other two sets of rankings for each story consisted of four items, the intrapersonal sets had five items because they included a nonsensical statement. The reasons for these omissions need to be explored by interviewing respondents after they take the test. Participants may have become frustrated at having to rank order five items instead of four, or the items themselves may have been problematic. Another potential issue with this set of rankings was the presence of a nonsensical item. From comments that several respondents provided in writing and verbally, there was some frustration associated with these items because of the polysyllabic words that they contain. One public school teacher circled these words throughout the measure and wrote a note to the researcher that a personality test should not include such uncommon words as “obsequiousness” and “toadyism.” In future revisions of RED, the nonsensical items may need to be toned down so that they do not intimidate or frustrate the reader.

Another issue that arose was that there were differences in the ratings of responses from one story to the next. Undergraduate students typically completed the measure in 35 to 45 minutes, and the order in which each story was presented did not vary. Therefore the order of presentation may have influenced responses, but it is not possible from this study to separate differences in item sets versus order effects.

Criterion-Related Validity. Although Study 2 assessed the construct validity of RED by comparing scores to a test that measures a similar construct, it did not explore criterion-related validity by comparing RED to another measure of the same construct. Future research should include subject-object interviews in order to establish criterion-related validity and to empirically develop a scoring formula for RED.

Discriminant Validity. Study 2 did not highly support Hypothesis 2, which stated that RED would not correlate with a measure of dispositional optimism, thereby establishing discriminant validity. However, only one of the four stage level scales moderately correlated with the LOT-R after controlling for age and education level, so dispositional optimism clearly had a weaker relationship with RED than the DIT, age, or educational level.

In future assessments, another measure should be tested with RED to fully establish discriminant validity, and one variable that should be considered for future research is vocational interest. For several decades, vocational counselors and researchers have utilized Holland's (1973, 1985) typology of vocational interests and work environments, demonstrating that a good fit between vocational interest and work environment is associated with job satisfaction and career stability (Holland, 1996). In recent years researchers have begun to pay greater attention to Holland's claim that vocational interest is an "expression of personality" (Holland, 1973, p.7), and have demonstrated correlations between vocational interest and personality traits that include extraversion, openness to experience, and conscientiousness (e.g. Barrick, Mount, & Gupta, 2003).

In his description of the mental demands of the modern workplace, Kegan never indicated that certain levels of meaning-making are associated with particular career interests. On the contrary, he makes it clear that our level of meaning-making has an impact on our ability to cope with the demands we face in all facets of life, both personal and professional. Within the workplace, our level of meaning-making influences our success in being able to cope with the responsibility and complexity present in our job *level*, whether we work as accountants, engineers, sales persons, psychologists, or housekeepers.

There is no published evidence that Kegan's stages would correlate with vocational interests. Furthermore, research has demonstrated that vocational interests remain fairly stable throughout the adult lifespan (Holland, 1996), indicating that our interests are not likely to change as we evolve from one stage of meaning-making to the next. For these reasons, establishing discriminant validity with a measure of vocational interest would appear to be a particularly apt choice.

In summary, the results of this initial research provided some support for the validity of RED as an assessment of an individual's stage of meaning-making. Continued development of the measure should include reaction statement revisions based upon respondent narratives, simplifying the reading level for stage-relevant and nonsensical reaction statements, and comparisons of different scaling methods and combinations of scales. In the present study, validity coefficients were highest for the ranked data. However, the effect that the rating exercise may have had on reaction statement rankings is not known. Different combinations of rankings and ratings should be tested to identify

a scaling method that will provide the necessary information to calculate valid scores without overburdening respondents with unnecessary tasks.

Research on the measure should continue with larger and more diverse samples and the inclusion of the subject-object interview score as a criterion variable. This research assessed the extent to which stage level scores from the RED measure correlated with stage level scores from the measure of a similar construct, the DIT. Future research to establish criterion-related validity should attempt to derive a scoring method that will estimate an individual's subject-object interview score within an acceptable range.

REFERENCES

- Affleck, G., Tennen, H., Zautra, A., Urrows, S., Abeles, M., & Karoly, P. (2001). Women's pursuit of personal goals in daily life with fibromyalgia: A value-expectancy analysis. *Journal of Counseling and Clinical Psychology, 69*, 587-596.
- Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *The Public Opinion Quarterly, 49*, 535-552.
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*, 45-74.
- Bartone, P., Bullis, R. C., Lewis, P., Forsyth, G. B., & Snook, S. (2001). *Psychological development and leader performance in West Point cadets*. Unpublished manuscript, United States Military Academy, Auburn University.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology, 69*, 25-39.
- Basseches, M. (1984). *Dialectical thinking and adult development*. Norwood, NJ: Ablex.
- Bellenger, B. L. (1999). A constructive-developmental examination of the propensity to engage in social loafing. *Dissertation Abstracts International, 60* (11), 5818B. (UMI No. 9949574).

- Brissette, I., Scheier, M. F., & Carver, C. S. (2002). The role of optimism in social network development, coping, and psychological adjustment during a life transition. *Journal of Personality and Social Psychology, 82*, 102-111.
- Collins, D. B., & Holton, E. F. (2004). The effectiveness of managerial leadership development programs: A meta-analysis of studies from 1982 to 2001. *Human Resource Development Quarterly, 15*, 217-248.
- Fejfar, M. C., & Hoyle, R. H. (2000). Effect of private self-awareness on negative affect and self-referent attribution: A quantitative review. *Personality and Social Psychology Review, 4*, 132-142.
- Fenigstein, A., & Levine, M. P. (1984). Self-attention, concept activation, and the causal self. *Journal of Experimental Social Psychology, 20*, 231-245.
- Forsyth, G. B., Snook, S., Lewis, P., & Bartone, P. (2002). Making sense of officership: Developing a professional identity for 21st century army officers. In D. M. Snider, G. L. Watkins, & L. J. Matthews (Eds.), *The future of the Army profession* (pp. 357-378). New York: McGraw Hill.
- Hamilton, J. C., & Shuminsky, T. R. (1990). Self-awareness mediates the relationship between serial position and item reliability. *Journal of Personality and Social Psychology, 59*, 1301-1307.
- Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliff, NJ: Prentice Hall.
- Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliff, NJ: Prentice Hall.

- Holland, J. L. (1996). Exploring careers with a typology: What we have learned and some new directions. *American Psychologist*, 51, 397-406.
- Jones, J. (1991). Assessing privacy invasiveness of psychological test items: Job-relevant versus clinical measures of integrity. *Journal of Business and Psychology*, 5, 531-535.
- Kegan, R. (1982). *The evolving self*. Cambridge, MA: Harvard University Press.
- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- Kegan, R., Lahey, L., & Souvaine, E. (1999). From taxonomy to ontogeny: Thoughts on Loevinger's theory in relation to subject-object psychology. In P. M. Westenberg, A. Blasi, & L. D. Cohn (Eds.), *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development* (pp. 39-56). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kegan, R., & Lahey, L. L. (2001a). *How the way we talk can change the way we work: Seven languages for transformation*. San Francisco, CA: Jossey-Bass.
- Kegan, R., & Lahey, L. L. (2001b). The real reason people won't change. *Harvard Business Review*, 79, 85-92.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally and Company.
- Kohlberg, L. (1981). *The meaning and measurement of moral development*. Worcester, MA: Clark University Press.

- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages*. San Francisco: Harper & Row.
- Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, 55, 312-320.
- Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings and the measurement of values. *The Public Opinion Quarterly*, 52, 526-538.
- Kuhnert, K. W., & Lewis, P. (1987). Transactional and transformational leadership: A constructive/developmental analysis. *Academy of Management Review*, 12, 648-657.
- Lahey, L. (1986). Males' and females' construction of conflict in work and love. *Dissertation Abstracts International*, 47(11), 4027B. (UMI No. 8704575).
- Lahey, L., Souvaine, E., Kegan, R., Goodman, R., & Felix, S. (1988). *A guide to the subject-object interview: Its administration and interpretation*. Cambridge, MA: Harvard Graduate School of Education.
- Laske, O. E. (2000). Foundations of scholarly consulting: The Developmental Structure/Process Tool (DSPT). *Consulting Psychology Journal: Practice and Research*, 52, 178-200.
- Laske, O. E. (November, 2004). *Looking for patterns in clients' developmental-behavioral dance with coaches*. Paper presented at the Second ICF Coaching Research Symposium, Quebec, Canada. Retrieved September 3, 2004, from [http://www. interdevelopmentals.org/Patterns%20in%20Clients%20Dance.pdf](http://www.interdevelopmentals.org/Patterns%20in%20Clients%20Dance.pdf)

- Leonard, H. S. (2003). Leadership development for the postindustrial, postmodern information age. *Consulting Psychology Journal: Practice and Research*, 55, 3-14.
- Lewis, P., Bartone, P., Forsythe, G. B., Bullis, C., Sweeney, P., & Snook, S. (2005). Identity development during the college years: Findings from the West Point longitudinal study. *Journal of College Student Development*, 46, 357-373.
- Lombardo, M. M., & Eichinger, R. W. (1989). *Eighty-eight assignments for development in place*. Report no. 136, Center for Creative Leadership, Greensboro, NC.
- Malcolm, J., & Ng, S. H. (1989). Relationship of self-awareness to cheating on an external standard of competence. *The Journal of Social Psychology*, 129, 391-395.
- McLean, L. A., Harvey, D. H. P., Pallant, J. F., Barlett, J. R., & Mutimer, K. L. A. (2004). Adjustment of mothers of children with obstetrical brachial plexus injuries: Testing a risk and resistance model. *Rehabilitation Psychology*, 49, 233-240.
- Merron, K., Fisher, D., & Torbert, W. R. (1987). Meaning-making and management action. *Group & Organization Studies*, 12, 274-286.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Ohlott, P. T., Ruderman, M. N., & McCauley, C. D. (1994). Gender differences in managers' developmental job experiences. *Academy of Management Journal*, 37, 46-67.
- Palus, C. J., Horth, D. M., Selvin, A. M., & Pulley, M. L. (2003). Exploration for development: Developing leadership by making shared sense of complex challenges. *Consulting Psychology Journal: Practice and Research*, 55, 26-40.

- Pratt, M. W., Diessner, R., Hunsberger, B., Pancer, S., & Savoy, K. (1991). Four pathways in the analysis of adult development and aging: Comparing analyses of reasoning about personal-life dilemmas. *Psychology and Aging, 6*, 666-675.
- Pyszczynski, T., Hamilton, J. C., & Herring, F. H. (1989). Depression, self-focused attention, and the negative memory bias. *Journal of Personality and Social Psychology, 57*, 351-357.
- Pyszczynski, T., Holt, K., and Greenberg, J. (1987). Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of Personality and Social Psychology, 52*, 994-1001.
- Rest, J. (1975). Longitudinal study of the *Defining Issues Test*: A strategy for analyzing developmental change. *Developmental Psychology, 11*, 738-748.
- Rest, J. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.
- Rest, J. (1990). *Manual for the Defining Issues Test* (3rd ed.). Minneapolis, MN: Center for the Study of Ethical Development.
- Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999a). A neo-Kohlbergian approach: The DIT and schema theory. *Educational Psychology Review, 11*, 291-324.
- Rest, J., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999b). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Rest, J., Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*, 5-28.

- Rest, J., Thoma, S., Narvaez, D., & Bebeau, M. J. (1997). Alchemy and beyond: Indexing the *Defining Issues Test*. *Journal of Educational Psychology, 89*, 498-507.
- Rooke, D., & Torbert, W. R. (1998). Organizational transformation as a function of CEO's developmental stage. *Organization Development Journal, 16*, 11-28.
- Rosse, J. G., Miller, J. L., & Stecher, M. D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology, 79*, 987-992.
- Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology, 64*, 219-238.
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4*, 219-247.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety), self-mastery, and self-esteem: A reevaluation of the *Life Orientation Test*. *Journal of Personality and Social Psychology, 67*, 1063-1078.
- Scheier, M. F., Matthews, K. A., Owens, J. F., Magovern, G. J., Lefebvre, R. C., Abbot, R. A., & Carver, C. S. (1989). Dispositional optimism and recovery from coronary artery bypass surgery: The beneficial effects on physical and psychological well-being. *Journal of Personality and Social Psychology, 57*, 1024-1040.

- Scheier, M. F., Weintraub, J. K., & Carver, C. S. (1986). Coping with stress: Divergent strategies of optimists and pessimists. *Journal of Personality and Social Psychology, 51*, 1257-1264.
- Schuler, H. (1993). Social validity of selection situations: A concept and some empirical results. In H. Schuler, J. L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 11-26). Hillsdale, NJ: Erlbaum.
- Senge, P. M. (2005). Missing the boat on leadership. *Leader to Leader, Fall(38)*, 28-30.
- Steginga, S. K., & Occhipinti, S. (2006). Dispositional optimism as a predictor of men's decision-related distress after localized prostate cancer. *Health Psychology, 25*, 135-143.

APPENDICES

APPENDIX A

Item Rating Form for the Reactions to Everyday Dilemmas Meaning-making Assessment

Reactions to Everyday Dilemmas is a scenario-based measure that asks respondents to indicate how they would react in each of five hypothetical situations. Participants are presented with five scenarios, followed by a series of items that are intended to reflect the meaning-making for Kegan's (1982, 1994) stages 2, 3, 4, or 5. In addition, the measure includes "high sounding" items (statements that sound sophisticated, but are essentially meaningless) to identify respondents who are not attending to the meaning of each statement.

This rating form has been developed to explore how well the measure's response items represent meaning-making associated with each stage:

- The form presents each scenario followed by thirteen response items. In the far left column, please indicate which meaning-making stage each item best reflects (2, 3, 4, or 5).
- If the item appears to be nonsensical, then write "HS" for "high sounding."
- If the item seems to reflect more than one stage, or does not appear to represent any particular stage of meaning-making (and is also not nonsensical), then place a question mark (?) in the box.
- Any comments you would like to provide in the far right column will be greatly appreciated.

For your reference, a brief description of each meaning-making stage is provided below.

Stage 2: Imperial. Individuals operating from the imperial stage are subject to personal interests, agendas, and role expectations. "Objects" are the impulses (e.g., the need for immediate gratification) to which one was subject in the previous stage. At Stage 2, one is capable of understanding and considering more than one perspective. However, one is unable to consider these viewpoints simultaneously and integrate them to generate a viewpoint that is a co-construction of one's own perspective and the perspective of someone else. Instead, one views other's actions and perspectives in light of the *potential consequences* these may have for one's own goals or agenda.

Stage 3: Interpersonal. The Stage 3 individual is subject to shared meaning, mutuality, social ideals and self-consciousness. As one transitions from Stage 2 to Stage 3, one develops the capacity to internalize other perspectives, and to develop a viewpoint that takes

multiple perspectives into account simultaneously. In other words, at Stage 3 the individual is capable of holding an internal conversation in which he or she considers how they feel, how others feel, and how they feel as a result of both perspectives. At Stage 3, one's self-concept is a co-construction of multiple perspectives – one's own opinion and the opinions of others.

Stage 4: Institutional. At Stage 4, individuals construct a self-authored system of values and standards that is used to reflect upon shared meaning. At this point, one's self-concept is no longer constructed with the opinions of significant others or societal ideals. As a result, Stage 4 individuals enjoy a psychological independence in which they recognize that their values and standards may differ from those of others (or from society). This makes it psychologically possible for Stage 4 individuals to grant others and themselves the freedom to possess and apply different standards.

Stage 5: Inter-Individual. At Stage 5, individuals are subject to what Kegan refers to as an “interpenetration of systems” (1982). At this stage of meaning-making one becomes open to considering the truths or value systems of others in a manner that allows one to comfortably recognize new values or truths for oneself that we had not allowed ourselves to consider before. According to Kegan (1994), Stage 4 individuals are able to comfortably recognize and visit opposing viewpoints as “tourists.” At Stage 5, relationships with those who hold opposing viewpoints are potentially transformational.

APPENDIX B

Reactions to Everyday Dilemmas Assessment for Study 1 (Instructions and Story A)

Welcome to the R.E.D. Inventory

Welcome to the R.E.D. Inventory. On the following pages you will be asked to read five brief stories and answer questions about how you would react to each of the situations described. These stories have been adapted from real situations that people have experienced in their work, school and social lives. While you may never have experienced situations identical to those depicted here, please try to imagine yourself in each of the stories that you will read, and think about how these experiences would make you feel.

This is the first version of R.E.D. that we have tested. Your responses to this inventory will help us identify content and formatting changes for future versions of the inventory. Please take your time and carefully consider what your reactions would be to each of the stories.

How to fill out the inventory:

There are 2 pages for each story and its questions. On the first page you will see the story (1 or 2 paragraphs) and a question that asks you to describe your reactions to the story in your own words.

On the second page you will be asked to answer 3 questions. Each question will have you rank order a set of statements based on how well you think each statement fits the underlying issues in the story or your reactions to the situation described in the story.

For each ranking question, you will see a list of statements with brackets to the left of each statement, similar to the table shown below. Please rank order the statements by writing a number next to each statement to indicate the statement's ranking.

Example:

The box shown below lists four possible reactions to a story. The numbers written on the left indicate that this respondent ranked the first statement as 3rd, the second statement was ranked 1st, the third statement was ranked 4th, and the fourth statement was ranked 2nd.

Please read all 4 or 5 statements for each question before writing any rankings.

- | | |
|-------|--|
| [3] | Has this situation changed the quality of my relationship with my supervisor? |
| [1] | This situation has presented new challenges for all of us. |
| [4] | It isn't clear exactly what the issues are for my team members. |
| [2] | This project will not work unless we both adopt our team roles, and work to meet the expectations we have of each other. |

To begin the inventory, please continue to the next page.

Story A:

Imagine that you have been working in a particular job position for almost two years. After receiving excellent performance reviews, you were recently promoted to a higher position that placed you in charge of several co-workers. Since your promotion, you have noticed that your co-workers have changed in their attitudes toward you. They are sometimes slow to respond to your requests, and have occasionally bypassed your authority by going to your boss with work issues.

Please take a moment to imagine yourself in the situation described in **Story A** and answer the 2 questions shown below. It might be helpful to think of a similar situation that you have actually experienced in real life

How would you feel in this situation? Why would you feel that way?

Story A:

1. Rank order the following issues in terms of importance from *Most Significant Issue (1)* to *Least Significant Issue (4)*.

- Has the promotion altered or ruptured the quality of my relationship with my former coworkers?
 - This situation may be new territory for all of us.
 - While it is clear that things are not working well here, it isn't clear exactly what the issues are for my former coworkers.
 - This employee-supervisor relationship cannot work unless we both adopt our respective roles, and work to meet the expectations we have of each other.
-

2. Consider how *you would feel* in this situation and rank order the statements from *Your Most Likely Reaction (1)* to *Your Least Likely Reaction (4)*.

- I could understand that my former co-workers might feel some resentment towards me. But ultimately, I'm not responsible for how they feel about being supervised by me. They need to take some responsibility for getting past their resentment, and I might be disappointed in them if they could not do so.
 - I could understand that they might resent my promotion. However, over time I would hope to demonstrate to them that coming to me first, as their supervisor, would not go unrewarded.
 - I would start to wonder whether or not I have misunderstood how my employees view my recent promotion. It's possible that it's not the attitudes of my employees that have changed, but rather my own perspective that has been transformed as a result of my promotion.
 - If my subordinates do not respect me as their supervisor, then I would find it difficult to feel good about myself until I was able to change the way they feel about me.
-

3. Consider how *you would feel* in this situation and rank order the statements from *Your Most Likely Reaction (1)* to *Your Least Likely Reaction (5)*.

- I should have realized that my supervisory authority is not a function of acting transformationally toward the individuals under my supervision. Instead, it must be based on a clearly established hierarchy of tasks that my employees have failed to comprehend.
- I would feel bad to think that I might not be living up to the expectations of the company and my boss. Even if no one else was aware of the problems I was having with my employees, I would feel uncomfortable knowing how disappointed my boss and my peers would be if they knew about the problems I was having.
- If I can't get my subordinates on board, my boss might conclude that I'm not doing a good job and I could be reprimanded and/or might not be offered other promotional opportunities in the future.
- If things were not working out between my employees and myself, then I would worry about whether or not I had been consistently and effectively applying the principles of supervision that I most highly value. But I might also be concerned about their level of professionalism.
- It would be important for me to recognize that there is more than one way to experience this situation. Perhaps I have hesitated to identify with my employees' concerns, because I would not want to see that there is still an "employee" in me as well as a "supervisor".

APPENDIX C

Reactions to Everyday Dilemmas Assessment for Study 2 (Instructions and Story A)

Welcome to the R.E.D. Inventory

Welcome to the R.E.D. Inventory. On the following pages you will be asked to read five brief stories and answer questions about how you would react to each of the situations described. These stories have been adapted from real situations that people have experienced in their work, school and social lives. While you may never have experienced situations identical to those depicted here, please try to imagine yourself in each of the stories that you will read, and think about how these experiences would make you feel.

Please take your time and carefully consider what your reactions would be to each of the stories.

How to fill out the inventory:

There are 2 pages for each story and its questions. On the first page you will see the story (1 or 2 paragraphs) and a question that asks you to describe your reactions to the story in your own words.

Next, you will be asked to rate and rank a set of statements that describe issues that might be important to you if you were in the situation described in the story.

It is likely that the lists of statements will not always include reactions that are exactly how you would feel in each situation. So even if you “rate” each statement as being low in importance, try to “rank” the statements based upon their relative importance compared to each other, even though the list may not include *your* most likely reaction.

Here is an example

Example Story

Imagine that you work in an office where all of the employees have had good working relationships with one another. However, the company has recently experienced a decrease in revenue and you have heard rumors that your office might have to layoff a couple of employees. Your supervisor suspects that one of your colleagues has been wasting a lot of time by surfing the Internet and writing personal e-mail on her office computer during company time. However, the supervisor has no proof of this. The supervisor has told you to spy on your co-worker so that they can catch her using the computer for personal business and fire her.

First, write a brief description of how you would feel in this situation, as shown below:

How would you feel in this situation? Why would you feel that way?

I would feel guilty about spying on my co-worker because I wouldn't want her to know that I helped to get her fired. I would think that it was unfair for my supervisor to put me in this situation without caring about how my co-workers might react if they knew I was doing this.

Next, you will be asked to rate and rank three sets of statements. In our example below, we only show one set of statements. However, the format of the ratings and rankings shown below are the same as in the actual test.

Read over each statement and rate each statement by placing an "X" in one of the four left-hand columns. You can use the same rating more than once. So if all of the statements would be very significant for you, or if they describe a reaction that you would be very likely to experience in the situation described, then you can place an X under "very" for all of the statements.

Next, you are asked to rank the 4 or 5 statements that are listed in each table by writing a number from 1 to 4 (or 5) next to each statement. Please do not give two statements the same ranking, even if your "ratings" for them are the same.

- Think about how significant each of the issues listed in the table below would be for you if you were in this situation, and rate each issue on a scale from **very significant** to **not at all significant** by marking X's in the appropriate columns.
- Next, rank order the issues in terms of importance from **Most Significant Issue (1)** to **Least Significant Issue (4)** by writing a number from 1 to 4 in the Ranking column.

2. Ratings				3. Rankings 1=Most Significant Issue	
Very	Somewhat	Not very	Not at all		
X				2	a. This is something my supervisor should be doing and it is not a part of my job.
X				1	b. My supervisor should know better than to ask me to do something that I may not be comfortable with.
	X			3	c. I am not sure if my coworker's offenses warrant this type of action.
		X		4	d. I don't consider myself one who would spy on my coworkers, but it seems as though I have to anyway.

You will be asked to provide ratings and rankings for three groups of times. The final task for each story will be to select the top 4 items out of all the ones listed on the page. In a table like the one shown below, select the 4 statements that you were best able to identify with by circling the letters that correspond to each of your preferred statements. Your four preferred statements may come from all three groups, or you might decide to select statements from only one or two groups.

- From all of the items listed above (statements a through m), select the top 4 that you most identify with by circling the letters below that correspond to your four most identifiable issues:

Most identifiable statement	a	(b)	c	d	e	f	g	h	i	j	k	l	m
Second most identifiable	(a)	b	c	d	e	f	g	h	i	j	k	l	m
Third most identifiable	a	b	(c)	d	e	f	g	h	i	j	k	l	m
Fourth most identifiable	a	b	c	(d)	e	f	g	h	i	j	k	l	m

To begin the inventory, please continue to the next page.

Story A:

Imagine that you have been working in a particular job position for almost two years. After receiving excellent performance reviews, you were recently promoted to a higher position that placed you in charge of several co-workers. Since your promotion, you have noticed that your co-workers have changed in their attitudes toward you. They are sometimes slow to respond to your requests, and have occasionally bypassed your authority by going to your boss with work issues.

1. Please take a moment to imagine yourself in the situation described in **Story A** and answer the questions shown below. It might be helpful to think of a similar situation that you have actually experienced in real life

How would you feel in this situation? Why would you feel that way?

2. Think about how significant each of the issues listed in the table below would be for you if you were in this situation, and rate each issue on a scale from **very significant** to **not at all significant** by marking X's in the appropriate columns.
3. Next, rank order the issues in terms of importance from **Most Significant Issue (1)** to **Least Significant Issue (4)** by writing a number from 1 to 4 in the Ranking column.

2. Ratings				3. Rankings 1=Most Significant Issue	
Very	Somewhat	Not very	Not at all		
					a. Has the promotion altered or ruptured the quality of my relationship with my former coworkers?
					b. This situation may be new territory for all of us.
					c. While it is clear that things are not working well here, it isn't clear exactly what the issues are for my former coworkers.
					d. This employee-supervisor relationship cannot work unless we both adopt our respective roles, and work to meet the expectations we have of each other.

4. Consider how *you would feel* in this situation and rate each of the following statements as describing a *very likely reaction, somewhat likely, not very likely, or not at all likely*.

5. Next, rank order the statements from *Your Most Likely Reaction (1)* to *Your Least Likely Reaction (4)*.

4. Ratings					5. Rankings 1=Most likely reaction	
Very	Some what	Not very	Not at all			
						e. I could understand that my former co-workers might feel some resentment towards me. But ultimately, I'm not responsible for how they feel about being supervised by me. They need to take some responsibility for getting past their resentment, and I might be disappointed in them if they could not do so.
						f. I could understand that they might resent my promotion. However, over time I would hope to demonstrate to them that coming to me first, as their supervisor, would not go unrewarded.
						g. I would start to wonder whether or not I have misunderstood how my employees view my recent promotion. It's possible that it's not the attitudes of my employees that have changed, but rather my own perspective that has been transformed as a result of my promotion.
						h. If my subordinates do not respect me as their supervisor, then I would find it difficult to feel good about myself until I was able to change the way they feel about me.

6. Consider how *you would feel* in this situation and rate each of the following statements as describing a *very likely reaction, somewhat likely, not very likely, or not at all likely*.

7. Next, rank order the statements from *Your Most Likely Reaction (1)* to *Your Least Likely Reaction (5)*.

6. Ratings					7. Rankings 1=Most likely reaction	
Very	Some what	Not very	Not at all			
						i. I should have realized that my supervisory authority is not a function of acting transformationally toward the individuals under my supervision. Instead, it must be based on a clearly established hierarchy of tasks that my employees have failed to comprehend.
						j. I would feel bad to think that I might not be living up to the expectations of the company and my boss. Even if no one else was aware of the problems I was having with my employees, I would feel uncomfortable knowing how disappointed my boss and my peers would be if they knew about the problems I was having.
						k. If I can't get my subordinates on board, my boss might conclude that I'm not doing a good job and I could be reprimanded and/or might not be offered other promotional opportunities in the future.
						l. If things were not working out between my employees and myself, then I would worry about whether or not I had been consistently and effectively applying the principles of supervision that I most highly value. But I might also be concerned about their level of professionalism.
						m. It would be important for me to recognize that there is more than one way to experience this situation. Perhaps I have hesitated to identify with my employees' concerns, because I would not want to see that there is still an "employee" in me as well as a "supervisor".

8. From all of the items listed above (statements a through m), select the top 4 that you most identify with by circling the letters below that correspond to your four most identifiable issues:

Most identifiable statement	a	b	c	d	e	f	g	h	i	j	k	l	m
Second most identifiable	a	b	c	d	e	f	g	h	i	j	k	l	m
Third most identifiable	a	b	c	d	e	f	g	h	i	j	k	l	m
Fourth most identifiable	a	b	c	d	e	f	g	h	i	j	k	l	m