Measuring Motivation Multidimensionally: Development of the Assessment of Individual Motives Questionnaire (AIM-Q)

Larry C. Bernard, Michael Mills, Leland Swenson and R. Patricia Walsh

Assessment 2008; 15; 16
DOI: 10.1177/1073191107306131

Measuring Motivation Multidimensionally

Development of the Assessment of Individual Motives–Questionnaire (AIM-Q)

Larry C. Bernard
Michael Mills
Leland Swenson
R. Patricia Walsh
Loyola Marymount University

We report the development of the Assessment of Individual Motives–Questionnaire (AIM-Q), a new instrument based on an evolutionary psychology theory of human motivation. It provides multitrait–multimethod (MTMM) assessment of individual differences on 15 motive scales. A total heterogeneous sample of N *= 1,251 participated in eight studies that evaluated the homogeneity, internal consistency, test–retest reliability, and MTMM convergent and discriminant validities of the AIM-Q's three methods. These studies generally support the overall strategy of assessing individual differences in multiple evolutionary-based motives with multiple methods. Additional validity studies are underway and, when validated further, the AIM-Q may offer a promising option for evolutionary psychologists and behavioral geneticists wanting to incorporate individual differences into their research but have had to use existing self-report measures of personality, which were not designed for such a purpose. It may also offer clinical and counseling psychologists an additional approach to personality measures for the prediction of behavior.*

Keywords: motivation; motives; measurement; assessment; evolutionary psychology; multitrait–multimethod

Individual differences have presented a problem for evolutionary psychologists and were neglected by them until recently (Buss & Greiling, 1999). There is now increasing interest in incorporating individual differences into evolutionary theory (Buss, 2004). For example, more use is being made of existing self-report measures of personality by evolutionary psychologists and behavioral geneticists (Ebstein, Benjamin, & Belmaker, 2003). Some of these existing measures include the NEO Personality Inventory–Revised (NEO PI-R; Costa & McCrae, 1992), Eysenck's Personality Inventory (Eysenck, 1952), and Cloninger's Tridimensional Personality Questionnaire (TPQ; Cloninger, 1987). However, these measures were originally derived from various personality theories and are not based on evolutionary psychology.

This poses a problem, because "personality" has no universally accepted definition among psychologists, although it is generally conceived as "distinctive patterns . . . that

The authors gratefully acknowledge the assistance of Aaron Lukaszewski, Charles Marriott, and Katherine Truitt with this research. Correspondence may be addressed to Larry C. Bernard, Psychology Department, Loyola Marymount University, One LMU Drive, Los Angeles, CA 90045; e-mail: lbernard@lmu.edu.

Assessment, Volume 15, No. 1, March 2008 16-35
DOI: 10.1177/1073191107306131

characterize each individual enduringly" (Mischel, 1999, p. 4). Such distinctive patterns may be the result of earlier adaptation and selection, as evolutionary psychologists would contend, but also may be because of other factors. For example, social psychology emphasizes the role of interpersonal factors in present behavior, developmental psychology emphasizes the role of early learning and experience, and Freudian psychology emphasizes the role of psychodynamic forces. In contrast, evolutionary psychology has a definite set of hypotheses about present behavior as a manifestation of earlier adaptation and selection (Buss, 2004). Therefore, individual differences constructs based on these evolutionary hypotheses should provide evolutionary psychologists and behavioral geneticists with measures that are more directly related to their phenomena of interest—that is, constructs as adaptations presumed to be *causing* enduring distinctive patterns of behavior.

We are attempting to bring the measurement of individual differences and evolutionary psychology together from the outset. Our goal is to produce an instrument that can be used to test evolutionary theory where it relates to human motivation. Such an instrument may also have utility in applied psychology, particularly with respect to predicting behavior. The instrument is the Assessment of Individual Motives–Questionnaire (AIM-Q) and it provides a multidimensional approach to measuring individual differences in the strengths of 15 human motives based on an evolutionary theory of human motivation.

THEORETICAL BACKGROUND

Bernard, Mills, Swenson, and Walsh (2005) recently presented a new evolutionary theory of human motivation. According to this theory, motivated behavior is purposeful behavior. Purposeful behavior is neither random nor simply reflexive. Purposeful behavior solves challenges to survival posed by what evolutionary psychologists call the "Environments of Evolutionary Adaptedness (EEA)," which "refers to the statistical composite of selection pressures that occurred during an adaptation's period of evolution responsible for producing that adaptation" (Buss, 2004, p. 40). Although, in general, EEAs are the distant past environments in which the genetic development of a species was shaped, an EEA is not a particular time or place, it is the selection forces responsible for shaping an adaptation. For example, with bipedal locomotion, the focus is on selection pressures operating about 4.4 million years ago. For example, with a psychological or behavioral adaptation such as conscience, the focus is on more recent selection pressures. Evolutionary psychology theory predicts that individuals will continue to emit these purposeful behavioral adaptations in the present (those behaviors that, in the aggregate, increased the likelihood of survival, now called "fitness," described below).

How is this possible? Bernard, Mills, et al. (2005) posit that purposeful human behavior is implemented by hypothesized neurocognitive structures that developed in EEAs. Much of motivated human behavior is presumed to be guided by these structures to solve the challenges of "individual fitness" and "inclusive fitness" (see Dawkins, 1982; Hamilton, 1964; Tooby & Cosmides, 1992). These challenges are historical, but not necessarily different from those still faced today.

Fitness

From an evolutionary perspective, individual fitness involves successfully overcoming two challenges: survival and reproduction. Individuals who engage in more self-protective behaviors such as aggression and/or submission may prolong their survival. Individuals who engage in more mating behavior such as sexual activity and status enhancement (through acquisition and display of resources) are more likely to reproduce.

Individual fitness is further enhanced by inclusive fitness (Hamilton, 1964; also see Dawkins, 1982), which involves successfully overcoming additional interpersonal challenges: forming and maintaining supportive relationships with partners, offspring, and kin; forming small (i.e., local), primarily reciprocal, coalitions with nonkin (strength in numbers); and facilitating large (i.e., national and international), primarily symbolic, coalitions that extend further the concept of kinship. For example, individuals who successfully demonstrate affection, altruism, and conscience, and who contribute to the general welfare, may provide a favorable social environment in which the likelihood of their offsprings' survival to reproductive age is increased.

Social Domains

The concept of inclusive fitness was a major step forward in the development of evolutionary psychology (Buss, 2004), and it was able to address such complex human social behaviors as grandparent investment in grandchildren and cooperative alliances between nonkin. This led logically to consideration of the social domains in which these newly appreciated, and more complex, inclusive fitness behaviors may operate. Bugental (2000) proposed that separate human neurohormonal systems evolved to overcome the challenges faced within different social domains. Kenrick, Li, and Butner (2003)

revised Bugental's proposed social domains by emphasizing their relevance to the rules and norms of human–social interaction.

Subsequently, Bernard, Mills, et al. (2005) noted that individuals face different challenges in different size social groups. They based their conclusion on Dunbar's (1993) evidence that the size of the human neocortex and the size of social groups in which humans interact may have coevolved and that the neurohormonal systems proposed by Bugental (2000) are similar to the neurocognitive systems described in Bernard, Mills, et al.'s (2005) evolutionary theory of human motivation. The social domains provide a framework for grouping purposeful behaviors within the AIM-Q and linking them to the size of the social group and the type of challenge they address.

We based the AIM-Q on Bernard, Mills, et al.'s (2005) refinements of both Bugental's (2000) and Kenrick, Li, and Butner's (2003) parsing of the social domains. Five social domains were ordered by the increasing size of the social systems to which they apply: (a) self-protection (individual systems); (b) mating (dyadic systems); (c) relationship Maintenance/Parental Care (small, mostly kin systems); (d) coalition formation (large, mostly nonkin systems); and (e) memetic (very large, heavily nonkin symbolic systems). This order also relates to the presumed chronological development of these social domains and the increasing cortical complexity necessary to mediate the purposeful behaviors within each domain.

Motives

Biological adaptations result in physical changes in organisms. Psychological adaptations are no less biological and also result in physical changes; however they are the changes that specifically affect the mental processes and behavior of organisms. Evolutionary psychologists consider each adaptation to have had a function in the EEA, to be universal, and to be discrete (Schmitt & Pilcher, 2004). We set out to identify the motivational adaptations necessary to solve major conceivable challenges to fitness within each of the five social domains. Each adaptation was hypothesized to result in a discrete set of behavioral tendencies that operated in the EEA. We use the term *motive* to refer to the psychological (neurocognitive) adaptations that mediate a set of behavioral tendencies. Motives are adaptational constructs that are presumed to guide behavior toward overcoming challenges within a specific social domain. They are independent, had discrete developmental histories, solved different challenges in the EEA, and are mediated by different neurocognitive adaptations.

We used a rational method and relied on evolutionary psychology theory to determine the motives. For the purposes of measurement, we operationalized the motives as individual differences in the *strength, desire, or concern about particular behaviors* that solve(d) problems of individual or inclusive fitness within each of the social domains. Three criteria were used to identify motives. A motive should (a) be related to one of the five social domains, (b) represent a cluster of purposeful behaviors that solved fitness challenges within that domain, and (c) be measurable as individual differences in terms of strength of interest, desire, or concern with those behaviors.

We used "classic ideas" in evolutionary theory (Mills, 2004) to help identify motives, our goal being to produce a comprehensive list of them. These classic ideas include the following: (a) natural selection—individual survival or what we have called "self-protection" (Darwin, 1859); (b) sexual selection—the evolution of specific traits that serve primarily to attract mates (Darwin, 1859); (c) inclusive fitness—altruism toward kin can have genetic payoffs (Hamilton, 1964); (d) parental investment theory—the sexes differ in investment to offspring and, thus, parental behavior (Trivers, 1972); (e) reciprocity—mutually beneficial reciprocal relationships with nonkin can be beneficial (Trivers, 1971); (f) generalized reciprocity—when reciprocity is "strong," that is enforced by "neutral" third party institutions, it can generalize to large groups of strangers (Gintis, 2000); and (g) memetic selection—genes are not the only replicators subject to evolutionary change, ideas can replicate and spread (Dawkins, 1976). We applied the principles of independence (discreteness) and parsimony in limiting the number of identified motives, but we also included several motives that may not meet the criterion of independence. This was done because this is a first attempt at defining the motives, and we desired to be more inclusive rather than overlook a potential motive. Therefore, we are prepared to drop or combine motives, depending on what the data reveal about their relationships. This process resulted in the identification of 15 putative motives, each related to a social domain. They are as follows: (a) Self-Protection Domain—Aggression, Curiosity, Health, Play, and Safety; (b) Mating Domain—Sex, and the "status motives" of Appearance, Material, Mental, and Physical; (c) Relationship Maintenance/Parental Care Domain—Affection; (d) Coalition Formation Domain—Altruism and Conscience; and (e) Memetic Domain—Legacy and Meaning. More detailed descriptions of the motives are provided in Table 1.

To avoid confusion, we point out that this is not a hierarchical model. The motives are independent and operate on "surface" behavior. Unlike hierarchical factor models common to personality measures, the social domains are not hierarchical factors influencing the motives. The social domains only describe specific realms of evolutionary challenges to which the motives are adapted. Because they are

TABLE 1
Descriptions of Assessment of Individual Motives–Questionnaire (AIM-Q) Scales

Social domain	*Motive*	*Operational Definition: Individual Differences in the Strength of Interest, Desire, or Concern With . . .*	*Evolutionary Challenge*
Self-protection	Aggression	Being physically dominant; combative; intimidating others	Protecting oneself (and later kin, and coalition members)
	Curiosity	Exploring new things, places, and situations; finding out about things, what they have to offer, and how they work	Understanding the physical environment to determine difficulties or find advantages to protect oneself, kin, and coalition members
	Safety	Being safe and secure in, and vigilant about, one's person and surroundings	Securing one's person, territory, and possessions (and later kin and coalition members) against hostile forces
	Play	Spontaneous or speculative activity; sportive, frivolous, mocking, or jesting behavior with others; free, unimpeded, stylish, or humorous interpersonal activity	Understanding the social environment, rules, reciprocity, and how people act, react, and interact through nonaggressive, mock situations
	Health	Remaining healthy and fit; improving one's health and fitness	Protecting one's physical integrity
Mating	Sex	Sexual activity; obtaining a desirable sexual partner	Genetic propagation
	Appearance	Improving one's physical appearance, grooming, and wardrobe; cosmetic attractiveness	Increasing status and desirability as a mate by appearing physically attractive
	Material	Acquiring assets; material competitiveness	Increasing status and desirability as a mate by accumulating material resources
	Mental	Developing one's knowledge, skills, and talents; nonathletic competition such as academics, games, arts, crafts, and hobbies	Increasing status and desirability as a mate by appearing smart and talented
	Physical	Developing one's physical strength or endurance; athletic competitiveness	Increasing status and desirability as a mate by appearing strong and dominant
Relationship, maintenance, and parental care	Affection	Tender attachment to others	Forming and maintaining cooperative alliances; maintaining relationships, and caring for young
Coalition formation	Altruism	Assisting others without self-benefit and possibly to one's personal detriment	Forming and maintaining cooperative alliances
	Conscience	Doing what is legally, morally, and ethically prescribed and avoiding what is proscribed; maintaining the traditions and rules of social interaction; reciprocity	Forming and maintaining cooperative alliances and maintaining relationships
Memetic	Legacy	The commonweal; making a lasting contribution, a better world for the next generation; leaving something of lasting value behind	Forming and maintaining broader (even symbolic) cooperative alliances to produce a general social–cultural community environment better suited to survival of kin and nonkin
	Meaning	Constructing a personal philosophy, meaning, or purpose for life	Explaining (rationalizing) one's existence and nonexistence

based strictly on evolutionary psychology theory, these motive constructs are different from personality constructs such as those in the Big Five model—neuroticism, extraversion, openness, conscientiousness, and agreeableness—and measured by such instruments as the NEO PI-R (Costa & McCrae, 1992). Each motive construct represents an independent adaptation in the EEA, whereas personality constructs such as those in the Big Five model were never conceived as such. Furthermore, each motive construct is intended to tap a single dimension of adaptation, whereas the NEO PI-R's Big Five constructs are multidimensional. For example, Conscientiousness subscales include competence, order, dutifulness, achievement striving, self-discipline, and deliberation.

In calling these motives, we do not intend to reify them. For example, we do not contend that there is a curiosity motive that is entirely genetically programmed. To do so would ignore the "strong interaction hypothesis" in evolutionary psychology, which holds that the environment also has a very powerful role in producing phenotype. Motives are adaptations in the neurocognitive mechanisms (defined in Bernard, Mills, et al., 2005) that mediate (propel) behavior. For example, curiosity is used for the sake of shorthand communication within the AIM-Q approach and represents

general exploratory, inquisitive behaviors (keeping in mind that the neurocognitive mechanism that mediates such behaviors has developed in interaction with environmental influences). Furthermore, we do not claim that any and all of an organisms' apparently exploratory movements would be motivated by a curiosity motive. We chose terms such as *curiosity* to describe the motives because they best represent and communicate what we believe to be an underlying *construct* about a category of purposeful behaviors familiar in human experience. We attempted to use terms that have relevance to evolutionary psychology theory rather than personality theory, although there is undoubtedly some overlap.

Finally, although the English language affords a rich variety of vocabulary, this can sometimes be a curse. Shades of meaning not found in other languages present difficulties when defining English terms in the sciences. For example, are these constructs "motives," "values," "goals," or "needs," to mention some related English language terms? The term motive has been used because the AIM-Q rests on an evolutionary psychology theory. Therefore, motive better connotes the hypothesized internal (neurocognitive) mechanisms believed to be implementing purposeful behavior. The other three related terms have also been used by motivational psychologists, but we believe they refer more to the challenges posed by each social domain, rather than the purposeful behaviors involved in overcoming the challenges.

For example, to call the conscience motive a value would ignore the "why question." Although most of traditional psychology has focused on the "how question" of behavior, that is its proximal causes, evolutionary psychology emphasizes the "why question" of behavior, its ultimate cause—why it exists at all. Evolutionary psychology emphasizes why a conscience motive should exist in the first place. One answer to the "why question" is that, when present and strong, the behaviors indicative of the conscience motive increased inclusive fitness in the EEA by allowing successful human interactions in large coalitions of nonkin. For that purpose, the conscience motive is defined as the propensity to engage in socially prescribed, and avoid socially proscribed behaviors and entering into reciprocal behavioral arrangements (Bernard, Mills, et al., 2005).

Likewise, to call the sex motive a goal is not incorrect, but would miss its importance as the means to the end—the evolutionary goal—of mating and reproduction. Terms such as *values* and *goals* are better suited to cognitive psychology, which deals more with proximal causes (see Bernard, Mills, et al., 2005, for a discussion of motivation and cognition). Finally, the term *need* is a closer approximation to our use of motive. However, need has sometimes been used to describe a deficiency state which observed behavior is presumed to correct (e.g., Murray, 1938). To call motives *needs* would conflict with the characterization of motives as "always on," as discussed in the next paragraph.

In conclusion, evolution does not produce perfect universal adaptations within a species; some individuals are better adapted to their environment than others. Therefore, individuals should vary in the strength of these motive adaptations and, if the motives are universal, this variation should allow motives to be measured as individual differences variables. The theory also predicts that motives are "always on," continuously organizing and directing behaviors. It is the strength of the organizing and directing of different motives that varies among individuals. Finally, the complexity and richness—even seemingly paradoxical nature—of observed purposeful human behavior results from different strengths of different motives acting independently and simultaneously. We maintain that much of what appears to be unpredictable (chaotic or spontaneous) in human behavior would become more predictable with knowledge of an individual's strengths of the various motives. The AIM-Q may provide a means of measuring the strength of motives.

GENERAL PROCEDURES USED TO DEVELOP THE AIM-Q

The AIM-Q was developed and refined in a 3-year, multistudy research project. We used the theoretical–rational method and followed Clark and Watson's (1995) and Smith and McCarthy's (1995) recommendations for objective scale development and refinement. They recommended an iterative process, several periods of item writing, and multiple samples, all of which were employed. The iterative process included several evaluations of unidimensionality (principal components analysis) alternating with internal consistency (coefficient α analysis). The results reported herein are from only the final stage of these analyses.

Clark and Watson (1995) recommended beginning with a clear conceptualization of the constructs and being overinclusive with items, which we did. We used several relatively large (N = 200-400) heterogeneous samples as well as some smaller samples that represent the target population. Clark and Watson also recommended that scale unidimensionality (defined as the degree to which a scale's items assess a single underlying construct) and validity be emphasized over internal consistency reliability. Therefore, we strove for mean interitem correlations in the range of .30 to .50, rather than trying to maximize coefficient α. We also followed Clark and Watson's suggestion to use either factor analysis or principal components analysis (which Cortina, 1993, recommends) to refine the unidimensionality

of scales. Principal components analysis was chosen to facilitate data reduction (Floyd & Widaman, 1995) and we compared it with the results of factor analysis. Both analyses yielded similar results. However, we were also mindful of Smith and McCarthy's argument that "reliance on item statistics is insufficient for determination of content homogeneity" (p. 305), and that some degree of judgment in determining items should also be used. Therefore, at this early point in development, we decided not to hold too strongly to the conventions of principal components analysis, by departing somewhat from Clark and Watson's recommendation to emphasize unidimensionality. Instead, we used principal components analysis for general guidance, and allowed items from some components with Eigenvalues < 1.0 to be included for analyses of internal consistency.

Multimethod Assessment

In developing the AIM-Q, we were also sensitive to the growing recognition of the need for multimethod assessment in psychology (Eid & Diener, 2006). Multimethod measurement facilitates the analysis of convergent and discriminant validity through recognition that a score reflects both its underlying psychological construct as well as systematic influences of the method used to derive the score. The AIM-Q's multimethod approach is represented in the incorporation of three different task–response formats, identified as "AIM1," "AIM2," and "AIM3." All formats are paper-based, but each requires a different task from the respondent. Therefore, these different formats are not intended to be alternative versions of the same instrument (although they may have utility as such).

The AIM1 consists of subtle items that presumably obscure the underlying motive constructs. The AIM2 consists of more obvious items typical of a traditional personality inventory, and respondents may be aware of the motive constructs being measured. The AIM3 consists of self- or other-endorsed ratings on each motive, and respondents would likely be aware of the motives being measured. The use of these subtle-obvious distinctions for the AIM-Q was influenced by Cattell, Horn, Sweney, and Radcliffe's (1964) strong recommendation to use disguised, subtle items when measuring motivation, but may also add method variance appropriate for a multimethod approach.

This study's primary objective was to develop three complimentary methods that may reliably and validly measure the 15 motives. If the motive constructs are robust across these methods, then they may be used to test the theory of motivation on which they are based. A secondary objective was to explore the concern of Cattell et al. (1964) for the need for disguised, subtle items in motivation research.

Samples and Procedures

Historically, the development of psychological instruments has been hampered by overreliance on samples of university students. Such samples were used because they were convenient and relatively economical, but they are fairly homogeneous. The use of student samples alone may restrict the amount of variance that a more heterogeneous sample would have. Therefore, we combined two techniques to obtain more heterogeneous samples that included both university students and community adults with an extensive range of ages from the beginning of development of the AIM-Q: (a) Internet administration and (b) acquaintance network sampling.

Internet administration. Internet administration involves obtaining responses over the Internet. All data were collected on (www.formsite.com), a commercial Web survey service. Respondents were assigned a unique code number that served as a password for the Internet site. The password had no information that could identify the respondent, thus ensuring anonymity. The password allowed partial course credit to be assigned to students without any connection to their responses or demographic data. Respondents were given the Internet address and asked to sign on at their convenience. A "welcome page" greeted respondents and was followed by the informed consent page. Participants who accepted the informed consent were then linked to the study instructions and began answering items. An additional advantage of the Internet procedure is that all items must be responded to, which eliminates incomplete data. After completing all items, participants answered a demographic questionnaire, were presented with a debriefing statement, and finally thanked.

Internet-based research may be no more risky than traditional research, and Internet samples tend to be relatively diverse with little problem of nonserious or repeat responders (Gosling, Vazire, Srivastava, & John, 2004; Kraut et al., 2004). There are, however, some limitations (Kraut et al., 2004). First is the problem of dropouts. People may dropout at any time with no immediate social consequences. However, inspection of the data revealed there were very few (<3%) terminated sessions and these could have resulted from either a dropped Internet connection or participants' voluntary discontinuation. This rate remained fairly constant across the studies. The second limitation is the potential for sampling bias. Although demographic differences between Internet users and nonusers have narrowed, Internet users are still more likely to be White, young, and have more children than the U.S. population as a whole (Kraut et al., 2004). We, therefore, urge caution when interpreting the generalizability of these findings until additional, more representative samples have been

TABLE 2
Study Number, AIM-Q Version, Sample Size, Age Range, Mean Age and Standard Deviation, Percentage of Women, Percentage of Nonstudents, and Percentage of Employed Nonstudents for All Studies

Study	*Version*	*Purpose*	N	*Age Range*	M *Age (*SD*)*	*Women*	*Nonstudents*	*Employed*
1	AIM1	Item Analysis 1	349	16 to 77	30.0 (12.5)	55%	68%	93%
2	AIM1	Item Analysis 2	206	18 to 78	35.8 (16.2)	56%	70%	90%
3	AIM1	Test–retest	16	20 to 71	34.3 (17.7)	56%	75%	94%
4	AIM2	Item analysis	419	15 to 78	29.5 (14.4)	51%	70%	93%
5	AIM2	Test–retest	22	18 to 24	20.7 (1.2)	91%	0%	N/A
6	AIM3	Test–retest	19	18 to 70	36.8 (16.5)	53%	74%	93%
7	AIM1,2,3	MTMM	116	18 to 59	30.7 (13.3)	57%	74%	92%
8	AIM1	Social desirability	38	18 to 22	18.9 (1.0)	68%	0%	N/A
	AIM2	Social desirability			50 selected randomly from the Study 4 sample			
	AIM3	Social desirability	66	18 to 64	32.6 (13.6)	65%	88%	92%

NOTE: N = 1,251. The n = 50 AIM2 Study 8 participants were counted in Study 4. AIM-Q = Assessment of Individual Motives–Questionnaire; MTMM = multitrait–multimethod matrix.

involved. However, we consider our community adult samples to better represent the adult population than university student samples alone.

Acquaintance network sampling. Acquaintance network sampling involves an initial sample of university students who participate themselves and then recruit acquaintances to participate. Students were instructed to recruit two nonstudents between 25 and 80 years of age. The specific advanced upper age limit was provided to encourage students to think of older adults who might participate. Students were requested to recruit friends (and given examples such as neighbors, teachers, coaches) or family (such as parents, grandparents, cousins). With Internet administration, recruits could be located anywhere—there were no geographic limits—and participate at any time convenient to them. Recruits were given passwords by their recruiters and directed to the appropriate Internet address. Recruits agreed to the same informed consent form, responded to the same items, and received the same debriefing as the initial student samples.

All student participants received partial course credit for their own participation and for the participation of their two recruits. In Studies 1 and 2, recruits were offered an incentive for participating by entry into a drawing for $100 money orders. Sign-up for the drawing was handled by a separate Web address to which recruits were directed after participating, so that personal information was not linked to respondents' demographic information or individual responses. The drawing may have encouraged recruits to repeat the process, although the length of the study should tend to discourage repeating it, and there was no evidence of repeaters. After Study 2, the drawings were dropped because: (a) about one third of recruits did not bother to enter the drawings and (b) feedback solicited from recruiters suggested that the recruits acted more out of social obligation than monetary incentive. It should be noted that, although this acquaintance network sampling procedure did successfully return a sample of nonstudent community adults with an extensive age range, as might be expected, on other demographic variables, the adult samples were not much different from their recruiters.

The remainder of this article describes the development of the AIM-Q through a series of studies that provide preliminary evidence of its reliability and internal validity. The same sampling and Internet administration procedures described above were used for all but one study. To facilitate comparison and reduce repetition, demographic data for each sample in each study are presented in Tables 2, 3, and 4. Sample sizes varied considerably depending on the purpose of the study. Larger samples were employed for factor and item analysis and smaller samples for other analyses such as test–retest reliability and social desirability. Most of the samples were fairly diverse ethnically (see Table 3) and in religious identification (see Table 4).

DEVELOPMENT OF THE AIM1

AIM1 items ask respondents to indicate what seems to them *the best use of a given resource such as time, effort, energy, or money*. People arguably perceive such resources as finite. Theoretically, resource allocation may reflect differences in the strength of motives. For example, a person who strongly agrees with this item—"one of the best uses of a free evening would be to spend it with somebody I loved in order to let them know how I felt"—is reporting that he or she would expend the resource *time* to give or receive *affection*. In their attempt to measure motivation, Cattell et al. (1964) called this allocation task

TABLE 3
Study Number, *N*, and Percentage of Participants' Within Each Ethnic Category for All Studies

	Ethnic Self-Identification Categories						
Study	*White Caucasian*	*Latino*	*Asian American*	*African American*	*Pacific Island Hawaiian*	*Mixed*	*Other*
1	217 (62%)	31 (9%)	29 (8%)	13 (4%)	—	29 (8%)	—
2	131 (64%)	35 (17%)	10 (5%)	12 (6%)	—	11 (5%)	—
3	14 (88%)	—	1 (6%)	1 (6%)	—	—	—
4	234 (56%)	79 (19%)	27 (6%)	20 (5%)	7 (2%)	27 (6%)	—
5	14 (64%)	1 (4%)	3 (13%)	2 (9%)	—	2 (9%)	—
6	13 (68%)	2 (10%)	1 (5%)	—	1 (5%)	1 (5%)	1 (5%)
7	59 (51%)	23 (20%)	17 (5%)	4 (3%)	2 (2%)	8 (7%)	1 (1%)
	21 (55%)	5 (13%)	5 (13%)	2 (5%)	2 (5%)	3 (8%)	—
8	Randomly selected sample from Study 4 (AIM2 Item Analysis)						
	21 (31%)	13 (19%)	8 (12%)	8 (12%)	6 (8%)	10 (15%)	—

NOTE: "Mixed" includes those participants whose parents were from two different ethnic groups.

TABLE 4
Study Number, *N*, and Percentage of Participants' Within Each Religious Category for All Studies

	Religious Self-Identification Categories						
Study	*Roman Catholic*	*Protestant*	*Agnostic Atheist*	*Jewish*	*Buddhist*	*Islam*	*Other*
1	164 (47%)	106 (30%)	44 (13%)	15 (4%)	4 (1%)	—	—
2	109 (53%)	52 (24%)	21 (10%)	10 (5%)	—	—	13 (6%)
3	5 (31%)	3 (19%)	4 (25%)	2 (12%)	—	—	2 (12%)
4	213 (51%)	114 (27%)	32 (8%)	10 (2%)	2 (1%)	8 (2%)	20 (5%)
5	12 (52%)	6 (26%)	2 (9%)	1 (4%)	—	1 (4%)	—
6	11 (58%)	3 (16%)	5 (26%)	—	—	—	—
7	66 (57%)	32 (28%)	8 (7%)	3 (3%)	—	3 (3%)	—
	16 (42%)	16 (42%)	3 (8%)	—	2 (5%)	—	1 (3%)
8	Randomly selected sample from Study 4 (AIM2 Item Analysis)						
	27 (40%)	26 (39%)	9 (13%)	2 (3%)	—	—	1 (2%)

"ends-for-means projection," suggesting that such items may tap how much an individual projects ends onto means (resources). The specific use to which a resource is put is presumed to reflect a motive. Respondents indicate their allocations of resources on a 4-point scale: 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, and 4 = *strongly agree*.

Study 1: AIM1 Item Analysis 1

Scale development and item writing was guided by Bernard, Mills, et al.'s (2005) theory of the evolutionary development of human motivation. The operational definitions of the 15 motives described in Table 1 provided guidelines for writing AIM1 items. Four psychologists from different areas (social, behavioral, evolutionary, and clinical psychology) and six advanced undergraduate psychology students wrote items. A first attempt at item writing produced 300 items (20 per scale). The items were then randomized and the same writers independently identified each of the 300 items with one of the 15 motives according to their operational definitions. Agreement of 80% or better for motive assignment was achieved for 190 items.

Some of the items with poorer interrater agreement were rewritten, and some entirely new items were added to the pool. A total of 306 surviving, edited, and new items were randomized and the same writers again independently sorted all of them onto one of the 15 motive scales. The 12 items with the highest rate of agreement (ranging from 80% to 100%) for each scale were selected for further analysis. Some scales had more than 12 items with similar high rates of agreement, in which case the items that would increase a scale's variety of behavioral

content were selected. Study 1 was a psychometric analysis of this initial 180-item pool, intended to determine if the interrater agreements of the item writers held up in an independent sample.

Procedure

The Internet administration and acquaintance network sampling methods and procedures were described in "General Procedures Used to Develop the AIM-Q." The Study 1 sample ($N = 349$) had a large range of ages, fairly diverse ethnic and religious self-identification, and about two thirds were community adults, (see Tables 2, 3, and 4 for more details). On completion of the study, community adult participants could enter a drawing for one of three $100 money orders.

Results and Discussion

Because Study 1 was preliminary, only general results from this analysis, but not the data, will be reported for the sake of brevity. A principal components (PC) analysis was performed to explore each scale's homogeneity with respect to its hypothesized underlying construct (Dunteman, 1989). Varimax rotation, which is customary, was not applied to this first solution, which was intended as an exploratory guide. We evaluated all components with Eigenvalues > 1.0, of which there were 2 to 4 for each scale.

The total variance accounted for by the first PCs of the 15 scales ranged from 21% to 41%. The number of items loading >.40 on the first PCs of each scale ranged from 5 to 12. Generally, only the first PCs were interpretable and all reflected the hypothesized underlying construct of each scale. As expected, the results of Study 1 indicated that, again, additional items needed to be written to increase the number of items per scale as well as the amount of variance accounted for by initial PCs.

Coefficient α was also computed for the items on the 1st PCs of every scale and ranged from .62 to .87. These values were judged adequate for a start, given the relatively short lengths of the 1st PCs.

Finally, intercorrelations were calculated for the 15 AIM1 scales to determine if the observed relationships between the scales were consistent with the theory. At this preliminary stage, the intercorrelations indicated that the scales were largely independent, as predicted by the theory, and that any correlations between them converged in a logical manner that supported the hypothesized underlying constructs of the scales. We were ready to proceed with additional item writing and editing in preparation for a second psychometric analysis.

Study 2: AIM1 Item Analysis 2

Study 2 incorporated all items with loadings ≥ .40 on the 1st PCs in Study 1 (AIM1 Item Analysis 1) with newly written items. The same 10 item writers evaluated the results of Study 1, specifically the underlying construct represented by each of the scales' first PCs, and used them as guides to write new items. Once again, scales were brought back up to 12 items each. With 15 scales, more than 12 items per scale would produce a lengthy instrument that might meet resistance by some respondents, particularly those who would be asked to complete additional instruments in subsequent validity studies. However, Smith and McCarthy (1995) caution against shortening scales too much because that may attenuate reliability (less of a problem) and validity (more of a problem because reliability may be preserved even though measurement of the target construct suffers from reduced coverage). We decided on 12 items per scale as a compromise. This would permit the removal of a few items from a scale, if necessary, without seriously abbreviating it.

Procedure

The Internet administration and acquaintance network sampling methods and procedures were described in "General Procedures Used to Develop the AIM-Q." The Study 2 sample ($N = 206$) had a large range of ages, fairly diverse ethnic and religious self-identification, and more than two thirds were community adults (see Tables 2, 3, and 4 for more details). On completion of the study, community adult participants could enter a drawing for a $100 money order.

Results and Discussion

Principal component analysis was again used to investigate the homogeneity of each scale. Varimax rotational criteria were employed to simplify identification of the components. Preliminary results suggested that two-component solutions were the most parsimonious and interpretable. Table 5 presents the summary results of the two-component varimax rotated analyses for each of the 15 scales with brief descriptions of each component. There are no agreed on criteria for how to proceed at this point (DeVellis, 1991; Dunteman, 1989; Spector, 1992; Thompson, 2004), but this is where having an evolutionary theory on which the scales are based is most helpful. Our goal was to produce factorially simple scales. However, because, unlike in Study 1 (AIM1 Item Analysis 1) where only the first PCs were identified with

TABLE 5
Study 2 AIM1 Principal Components Analysis Results

				Variance	
Scale	*Component*	*Label*	*Eigenvalue*	*%*	*Cumulative %*
Affection	1	Time spent obtaining	3.496	35.0	
	2	Money spent obtaining	1.117	11.2	46.2
Aggression	1	Passive-vicarious dominance	3.092	25.8	
	2	Active—revenge	1.910	15.9	41.7
Altruism	1	Active—service to others	2.449	22.3	
	2	Sharing resources	1.423	12.9	35.2
Appearance	1	Physical enhancement	3.889	35.4	
	2	Cosmetic surgery	1.300	11.8	47.2
Conscience	1	Doing right	2.629	26.3	
	2	Making amends	1.428	14.3	40.6
Curiosity	1	Exploration–exotic travel	3.848	32.1	
	2	Inquiring mind	1.200	10.0	42.1
Health	1	Active—improvement	3.490	34.9	
	2	Passive—information gathering	1.397	14.0	48.9
Legacy	1	Investing in future generations	3.421	28.5	
	2	Improving the future	1.423	11.9	40.4
Material	1	Flaunting wealth	4.125	37.5	
	2	Accumulating wealth	1.192	10.8	48.3
Meaning	1	Inquiring about meaning	4.494	40.8	
	2	Reflecting about meaning	1.131	10.3	51.1
Mental	1	Acquiring knowledge	4.021	36.6	
	2	Improving mental abilities	1.434	13.0	49.6
Physical	1	Enhancing physical stature	3.802	42.2	
	2	Enhancing strength	1.089	12.1	54.3
Play	1	Active frivolity	3.581	32.6	
	2	Passive game playing	1.227	11.2	43.8
Safety	1	Harm avoidance	3.523	32.0	
	2	Seeking security	1.249	11.4	43.4
Sex	1	Learning about sex	4.822	40.2	
	2	Sexual activity	1.187	9.9	50.1

NOTE: $N = 206$.

the motive constructs, for all scales in Study 2, both 1st and 2nd PC's were readily identifiable as related to the theory-predicted motive constructs. Therefore, we decided to include items from both components in the next stage of analysis. Together, both components accounted for between 40% and 54% of the total variance in each scales' items. We were able to readily identify the PCs, and each was given a brief descriptive label that is consistent with its motive's hypothesized underlying construct as indicated in Table 5.

We then used internal consistency analysis to indicate which individual items should be eliminated. Internal consistency was evaluated by computing coefficient α for each scale. After inspection of these values in conjunction with the principal components analyses, several scales' αs were increased by eliminating one or two items. The results at each scale's optimum α level are depicted in Table 6. Final scale lengths now range from 10 to 12 items and final αs range from .64 to .86. All AIM1 scales, except perhaps Altruism (α = .64) and

TABLE 6
Study 2 AIM1 Scales' Internal Consistency (coefficient α)

Scale	*Items*	*α*
Affection	11	.78
Aggression	12	.73
Altruism	11	.64
Appearance	11	.81
Conscience	10	.68
Curiosity	12	.80
Health	10	.76
Legacy	12	.76
Material	11	.83
Meaning	12	.85
Mental	11	.82
Physical	11	.84
Play	12	.77
Safety	11	.75
Sex	12	.86
Total Items	169	

NOTE: $N = 206$.

TABLE 7
Study 2 AIM1 Scale Intercorrelations

Scale	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1. Affection	—														
2. Aggression	–.26	—													
3. Altruism	.47	–.17	—												
4. Appearance	–.03	.20	–.16	—											
5. Conscience	.55	–.29	.50	–.15	—										
6. Curiosity	.42	–.02	.35	.10	.33	—									
7. Health	.39	–.06	.33	.00	.44	.41	—								
8. Legacy	.53	–.13	.64	.00	.48	.51	.43	—							
9. Material	–.19	.34	–.24	.56	–.36	–.06	–.18	–.16	—						
10. Meaning	.02	–.02	.31	–.01	.11	.32	.18	.31	–.08	—					
11. Mental	–.31	.48	–.16	.26	–.38	.06	–.15	–.15	.48	.31	—				
12. Physical	–.17	.46	–.12	.25	–.23	.00	.30	.30	.33	.12	.51	—			
13. Play	.48	.01	.25	.04	.28	.41	.26	.26	.02	.11	.11	.14	—		
14. Safety	.36	–.12	.15	.21	.34	.20	.32	.32	.11	–.14	–.20	–.08	.08	—	
15. Sex	–.05	.33	–.02	.19	–.20	.29	.08	.02	.32	.24	.40	.29	.33	–.21	—

NOTE: $N = 206$; all $|r| > .13$ are significant at $p < .05$; all $|r| > .18$ are significant at $p < .01$.

Conscience ($\alpha = .68$), were considered to have adequate internal consistency reliability.

Intercorrelations were calculated for the 15 AIM1 scales (see Table 7). These correlations represent relationships between the revised scales. Because of the relatively large sample size, weak correlations ($|r| = .10$) are significant, $p < .05$, so we interpreted only correlations for which $|r| \geq .31$, which accounts for ≥10% of the variance between scales. Out of 105 intercorrelations, 35 were greater than $|r| = .31$. All of them were in the moderate range and positive. The 70 (66%) intercorrelations of $|r| < .31$ suggests a fairly high degree of independence in the scales, as predicted by the theory, and the intercorrelations $|r| \geq .31$ generally converged in a manner supportive of the underlying constructs of the scales.

Study 3: AIM1 Test–Retest Reliability

Procedure

The AIM1 scales' test–retest reliability was assessed in a different, smaller sample ($N = 16$) of participants who took it on two occasions, one week apart. The Internet administration and acquaintance network sampling methods and procedures were the same as described in "General Procedures Used to Develop the AIM-Q." The Study 3 sample still had a large range of ages and fairly diverse religious self-identification, but there was less diverse ethnic identification. Three quarters of participants were community adults (see Tables 2, 3, and 4 for more details).

Results and Discussion

This procedure is intended to provide an estimate of temporal stability in a measure and it has the potential to indicate how much error there is in the AIM1 scales, provided the underlying constructs are themselves stable (DeVellis, 1991). Test–retest reliability can overestimate reliability because of carryover memory effects from the first to the second testing (Carmines & Zeller, 1979). However, the length of 169 items probably mediates against inflated reliabilities because of memory. The test–retest reliabilities for each of the AIM1 scales may be found in the column labeled "AIM1" under "Monotrait–Monomethod" in Table 8. (To avoid repetition and save space, the results from Studies 3 [AIM1 test–retest], 5 [AIM2 test–retest], 6 [AIM3 test–retest], and 7 [multitrait–multimethod (MTMM)] are presented together in Table 8.) With correlations between the two administrations ranging from $r = .80$ to .96 in this sample, AIM1 scales appear to have reasonably good test–retest reliability, or conversely, reasonably little error in this sample.

DEVELOPMENT OF THE AIM2

The AIM1 was developed to provide an indirect, disguised task through which to assess individual differences in the strength of the 15 motives and it is presumed to be less subject to the effects of bias or distortion (this remains to be demonstrated, of course, in subsequent research). The AIM2 was developed as a more direct task

TABLE 8
Studies 3 (AIM1Test–Retest), 5 (AIM2 Test–Retest), and 6 (AIM3 Test–Retest), Monotrait–Monomethod Correlation Coefficients (Test–Retest Reliabilities) and Study 7 (MTMM) Monotrait–Heteromethod Correlation Coefficients (Convergent Validities) of AIM-Q Motive Scales with their Summary Medians, Means, and Standard Deviations

	Monotrait–Monomethod			*Monotrait–Heteromethod*		
Motive	*AIM1*	*AIM2*	*AIM3*	*AIM1-AIM2*	*AIM1-AIM3*	*AIM2-AIM3*
Affection	.89	.69	.85	.70	.57	.68
Aggression	.92	.78	.94	.58	.36	.41
Altruism	.80	.88	.51	.64	.43	.39
Appearance	.84	.78	.91	.58	.62	.62
Conscience	.82	.73	.72	.62	.52	.62
Curiosity	.96	.77	.87	.53	.57	.56
Health	.80	.84	.80	.57	.49.	.60
Legacy	.90	.85	.76	.74	.54	.54
Material	.90	.83	.82	.52	.51	.65
Meaning	.94	.85	.77	.71	.73	.76
Mental	.89	.91	.75	.44	.51	.59
Physical	.92	.86	.81	.78	.68	.70
Play	.90	.73	.56	.59	.55	.58
Safety	.82	.88	.78	.56	.57	.66
Sex	.92	.91	.88	.57	.51	.69
Md	.90	.84	.80	.58	.54	.62
M	.88	.82	.78	.61	.54	.60

to assess the strength of the 15 motives. Its items are not disguised and resemble traditional measures of personality traits. AIM2 items ask respondent to endorse statements about interests, desires, wants, and likes that people may have as *true* or *untrue* of them using a 6-point scale: 1 = *completely untrue*, 2 = *mostly untrue*, 3 = *slightly untrue*, 4 = *slightly true*, 5 = *mostly true*, and 6 = *completely true*. Examples of AIM2 items include as follows: (a) I like sex, (b) I want to have plenty of money to do what I want, and (c) I am interested in developing affectionate attachments.

Study 4: AIM2 Item and Scale Analysis

Experience with the operational definitions of the 15 motive scales and with the constructs obtained through the item analyses of AIM1 facilitated the writing of items for the AIM2. The first author wrote 180 new items, 12 per scale, for the AIM2.

Procedure

The Internet administration and acquaintance network sampling methods and procedures were the same as described in "General Procedures Used to Develop the AIM-Q." The Study 4 sample ($N = 419$) had a large range of ages, fairly diverse religious and ethnic self-identification, and more than two thirds were community adults (see Tables 2, 3, and 4 for more details).

Results and Discussion

The same analytical procedures used in Study 2 (AIM1 Item Analysis 2) were repeated for the AIM2. Principal component analysis was used to investigate the homogeneity of each scale, with varimax rotational criteria employed to simplify identification of the components. Results suggested that two-component solutions were again the most parsimonious and interpretable. Table 9 presents the summary results of the two-component varimax rotated analyses for each of the 15 scales, with brief descriptions of each component. It is customary, but by no means obligatory, to include only PCs with Eigenvalues >1.0. All Eigenvalues of the AIM1's 15 scales' first two PCs were >1.0, whereas 7 of the AIM2's scales' second PCs had Eigenvalues <1.0. However, some items on the these PCs with Eigenvalues <1.0 had face-validity with respect to the underlying constructs and yet sampled from different content domains. Therefore, to be overinclusive, we decided to include all items from the AIM2's second PCs as well, and to allow any weaker items to be eliminated during the subsequent analysis of internal consistency.

Comparison of the descriptions of the AIM2 PCs in Table 9 with those of the AIM1 PCs in Table 5 suggests

TABLE 9
Study 4 AIM2 Principal Components Analysis Results

				Variance	
Scale	*Component*	*Label*	*Eigenvalue*	*%*	*Cumulative %*
Affection	1	Tender attachment to others	6.287	57.2	
	2	Displays of tenderness	0.797	7.2	64.4
Aggression	1	Intimidation of others	3.396	28.3	
	2	Not backing down	1.625	13.5	41.8
Altruism	1	Helping others	4.861	40.5	
	2	Helping anonymously	1.332	11.1	51.6
Appearance	1	Physical enhancement	5.853	53.2	
	2	Looking good	0.948	8.6	61.8
Conscience	1	Following rules	5.121	42.7	
	2	Doing right	1.519	12.6	55.3
Curiosity	1	Inquisitiveness	4.757	47.6	
	2	Exploring the new	0.941	9.4	57.0
Health	1	Active—improvement of health	5.703	51.8	
	2	Passive—wanting better health	1.030	9.4	61.2
Legacy	1	Leaving something positive behind	5.509	55.1	
	2	Improving the future	0.907	9.1	64.1
Material	1	Wealth and financial success	6.167	51.4	
	2	Money for status	1.039	8.7	60.1
Meaning	1	Meaning of life	6.380	53.2	
	2	Mysteries of the universe	0.985	8.2	61.4
Mental	1	Knowing more than others	4.298	39.1	
	2	Academic achievement	1.878	17.1	56.2
Physical	1	Physical domination	5.919	53.8	
	2	Physical enhancement	0.958	8.7	62.5
Play	1	Frivolity	5.319	48.4	
	2	Playful approach to life	0.944	8.6	57.0
Safety	1	Security	6.119	51.0	
	2	Harm avoidance	1.297	10.8	61.8
Sex	1	Sexual activity	7.717	64.3	
	2	Gratuitous sexual activity	1.002	8.4	72.7

NOTE: $N = 419$.

that the AIM1 and AIM2 Motive scales tap slightly, but not altogether, different domains. Most of the difference in the descriptions appears to result from the different AIM1 and AIM2 methods. For example, consider the affection motive. Its two AIM1 PCs are time spent obtaining affection and money spent obtaining affection. Its two AIM2 PCs are (interested in) tender attachments to others and (desiring) displays of affection. The difference in assessment method is clearly evident in the contrast between the AIM1 and AIM2 PCs. For all of the AIM2 scales, both components were readily identifiable as related to the same theory-predicted motive constructs as the AIM1 scales. Each AIM2 scale's two components accounted for between 42% and 73% of the total variance in the scales' items, somewhat more than was accounted for by the AIM1 scales' first two components.

Coefficient α was computed for each AIM2 scale. After inspection, several scales' αs were increased by eliminating one or two items. The results at each scale's optimum α level are depicted in Table 10. Final scale lengths now range from 10 to 12 items and final αs range from .76 to .95. Internal consistency in this sample appears to be slightly better for AIM2 versus AIM1.

Intercorrelations were calculated for the 15 AIM2 scales and are presented in Table 11. These correlations represent convergent and discriminant relationships between the scales. Because of the relatively large sample size, weak correlations ($|r| = .09$) are significant, $p < .05$. Therefore, interpretations were again based on relationships of $|r| = .31$, which account for at least 10% of the variance between scales. Out of 105 intercorrelations, 28 were greater than $r = .31$. All of them were in the moderate range and positive. The 77 (73%) intercorrelations of $r < .32$ suggests a slightly higher degree of independence between the scales of the AIM2 than the AIM1. The intercorrelations $|r| = .31$ generally converged in a manner supportive of the underlying constructs of the scales.

TABLE 10
Study 4 AIM2 Scales' Internal Consistency (coefficient α)

Scale	*Number of Items*	*α*
Affection	11	.91
Aggression	12	.76
Altruism	12	.86
Appearance	11	.90
Conscience	12	.86
Curiosity	10	.87
Health	11	.90
Legacy	10	.90
Material	12	.91
Meaning	12	.91
Mental	11	.84
Physical	11	.91
Play	11	.88
Safety	12	.91
Sex	11	.95
Total Items	169	

NOTE: $N = 419$.

Study 5: AIM2 Test–Retest Reliability

Procedure

The AIM2 scales' test–retest reliability was assessed in a different, smaller sample ($N = 22$) of participants who took it on two occasions, 1 week apart. The Internet administration procedures were the same as described in "General Procedures Used to Develop the AIM-Q." Because all Study 5 participants were university students, the Study 5 sample had a restricted range of ages and less diverse ethnic identification (see Tables 2, 3, and 4 for more details).

Results and Discussion

The test–retest reliabilities may be found in the column labeled AIM2 under "Monotrait–Monomethod" in Table 8. They are the correlations between the two administrations, and ranged from $r = .69$ to .91 in this sample. AIM2 scales would appear to have slightly lower test–retest reliability coefficients, or conversely, slightly more error than AIM1 scales. However, reliability estimates are affected by sample characteristics and should not be taken as fixed qualities of a specific scale (Reese, Kieffer, & Briggs, 2002). The samples used for the AIM1 and AIM2 test–retest reliability studies were demographically different—the former predominantly nonstudent adults and the latter all university students—and this may have affected the outcome. It is also possible that differences in item transparency—the difference in the methods themselves—may have contributed to differences in test–retest reliabilities. Further studies will be necessary to obtain additional estimates of test–retest reliability for the AIM1 and AIM2 scales in new samples and under different conditions. Only the convergence of such data can yield relatively accurate estimates of reliability. In the meantime, these test–retest reliability coefficients may be taken as preliminary estimates of measurement error in the scales.

DEVELOPMENT OF THE AIM3

The AIM1 and AIM2 both aggregate scores on multiple items. In contrast, the task of the AIM3 is to rate oneself (or another) on single, global descriptions of each of the 15 motives. There is evidence that the validity of very short self-rating scales may be as good as, if not better than, that of multiple item scales (Burisch, 1984a, 1984b, 1997). For example, single item self-ratings may account for equivalent amounts of variance in criterion ratings of some of the Big-Five personality domains when compared to the longer scales of the NEO PI-R (Bernard, Walsh, & Mills, 2005). There is even some evidence that single-item global self-ratings can yield greater criterion validity than aggregate scales (e.g., Burisch, 1984a, 1984b). Others, however, dispute that evidence, and argue that aggregate scales based on classical test theory are more reliable and valid (Paunonen & Jackson, 1985a, 1985b). This is an issue that can and should be settled empirically.

Therefore, the AIM3 was developed as a third method of measuring the 15 motives. The operational definitions of the motives (see Table 1) were used as the basis for each of the motive descriptions. The descriptions were slightly rephrased and the motives were described as *different goals or priorities people may have in their lives* rather than as motives, per se. (Note that the use of terms such as goals and priorities is done for ease of understanding on the part of respondents. In our discussions with participants, it has been easier for them to understand terms such as these, rather than terms related to evolutionary and motivational theory.) Respondents would then indicate how important each of the goals and priorities were to them using a 6-point scale: 1 = *no importance at all*, 2 = *a lot less than average importance*, 3 = *a little less than average importance*, 4 = *a little more than average importance*, 5 = *a lot more than average importance*, and 6 = *great importance*. An example of an AIM3 item is as follows: "taking actions to improve my health and avoid illness," for the health motive.

Single-item measures do present issues for psychometric analysis because they do not lend themselves to some of the standard methods of assessing reliability. Nevertheless, they can be subjected to test–retest reliability studies and

TABLE 11
Study 4 AIM2 Scales Intercorrelations

Scale	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1. Affection	—														
2. Aggression	.00	—													
3. Altruism	.38	–.09	—												
4. Appearance	.36	.23	.24	—											
5. Conscience	.27	–.16	.24	.20	—										
6. Curiosity	.35	.25	.44	.24	.24	—									
7. Health	.27	.06	.28	.40	.50	.28	—								
8. Legacy	.35	.05	.48	.29	.36	.48	.36	—							
9. Material	.19	.29	.21	.45	.10	.21	.18	.20	—						
10. Meaning	.29	.10	.46	.17	.24	.46	.16	.54	.04	—					
11. Mental	.20	.39	.36	.38	.15	.36	.21	.32	.40	.28	—				
12. Physical	.01	.43	–.11	.27	–.16	.09	.18	.00	.27	.02	.34	—			
13. Play	.44	.24	.39	.24	–.01	.39	.15	.14	.21	.16	.10	.18	—		
14. Safety	.23	–.24	.22	.23	.64	.00	.50	.30	.10	.13	.08	–.24	–.15	—	
15. Sex	.19	.30	.17	.20	–.25	.17	–.02	–.06	.16	–.04	.15	.24	.28	–.24	—

NOTE: $N = 419$; all $|r| > .09$ are significant at $p < .05$; all $|r| > .12$ are significant at $p < .01$.

the degree of relationship between the AIM3 and the AIM1 and AIM2 scales can also be examined.

Study 6: AIM 3 Test–Retest Reliability

Procedure

The AIM3 scales' test–retest reliability was assessed in a different sample ($N = 19$) of participants who took it on two occasions, 1 week apart. The Internet administration and acquaintance network sampling methods and procedures were the same as described in "General Procedures Used to Develop the AIM-Q." The Study 6 sample still had a large range of ages and fairly diverse religious self-identification, but there was less diverse ethnic identification. Nearly three quarters of participants were community adults (see Tables 2, 3, and 4 for more details).

Results and Discussion

The test–retest reliabilities may be found in the column labeled AIM3 under "Monotrait–Monomethod" in Table 8. Correlations between the two administrations ranged from $r = .51$ to .94 in this sample. Most of the single-item AIM3 scales have test–retest reliabilities similar to the lengthier AIM1 and AIM2 scales. This tends to support the position that single-item global ratings may be as reliable as aggregate scales. However, it is noteworthy that the test–retest reliabilities of the AIM3 Altruism ($r = .51$) and Play ($r = .56$) scales are considerably lower than those obtained with the AIM1 and AIM2. This mixed pattern of high and low test–retest reliabilities for single-item scales has been found with other instruments (Bernard, Walsh, et al., 2005) and suggests that some dimensions may not be suited to single-item assessment. An alternate explanation would be that some motives may be more likely to fluctuate more in reaction to time or external factors. As discussed, conclusions should be withheld pending further studies to obtain additional estimates of test-retest reliability in new samples and under different conditions as well as to investigate the reactivity of motives measured by AIM3 versus the aggregate methods.

CONVERGENT AND DISCRIMINANT VALIDITY STUDIES

Study 7: MTMM Analysis of AIM1, AIM2, and AIM3

The utility of the AIM-Qs multimethod approach depends on the degree to which its three methods reliably and validly measure the same 15 motives. Campbell and Fiske (1959) proposed the use of a MTMM matrix for the examination of convergent and discriminant validity. This method does have limitations, primarily in that it lacks objective criteria and statistical tests to evaluate the pattern of differences among correlations in the MTMM matrix (Schmitt, 2006). Still, the MTMM method has been influential in the construction and validation of new instruments and it does provide a standard method to present MTMM data (Schmitt, 2006). An MTMM matrix requires a single sample to complete all three AIM-Q methods.

Procedure

The Internet administration and acquaintance network sampling methods and procedures were described in "General Procedures Used to Develop the AIM-Q." The Study 7 sample ($N = 116$) had a large range of ages, fairly diverse ethnic and religious self-identification, and nearly three quarters were community adults (see Tables 2, 3, and 4for more details). The AIM1 and AIM2 were presented in counterbalanced order based on random assignment. The AIM3 followed both last because it describes the 15 motive traits.

Results and Discussion

Campbell and Fiske's (1959) recommended presentation of an MTMM structure matrix consists of three monomethod blocks—the correlations of each scale with itself and with all other scales measured by the *same* method—and three heteromethod blocks—the correlations of each scale with itself and with all other scales as measured by all pairs of *different* methods. When applied to the AIM-Q, the resulting MTMM structure matrix, at 45 by 45 cells, is too large to present in a readable table in standard journal format. To preserve space, we present summary statistics, that is *M* values, for each of the six MTMM blocks. (The complete MTMM matrix is available on request from the first author.)

The main diagonal in the MTMM is the MTMM correlation of a scale with itself, the test–retest reliabilities. These were obtained in Studies 3 (AIM1 test–retest), 5 (AIM2 test–retest), and 6 (AIM3 test–retest) and are found in columns labeled AIM1, AIM2, and AIM3 under MTMM in Table 8. The *M* test–retest reliabilities across the 15 scales found at the bottom of each column are as follows: .88 ($SD = .05$) for the AIM1, .82 ($SD = .07$) for the AIM2, and .78 ($SD = .12$) for the AIM3. These values provide a gauge of the overall reliabilities of the three AIM-Q measures and suggest that the AIM-Q scales have generally acceptable short-term stability.

We next consider the monotrait–heteromethod correlations. These values are found in columns labeled AIM1-AIM2, AIM1-AIM3, and AIM2-AIM3 under monotrait–heteromethod in Table 8. The *M* values of the correlations of a scale with itself for each pair of different methods across the 15 scales found at the bottom of each column are as follows: .61 ($SD = .09$) for the AIM1-AIM2, .54 ($SD = .09$) for the AIM1-AIM3, and .60 ($SD = .10$) for the AIM2-AIM3. When compared with the MTMM correlations in Table 8, these monotrait–heteromethod correlations provide a sense of the overall influence of method on the AIM-Q scales. Each scale's monotrait–heteromethod correlation should be moderate and smaller than its MTMM correlation. This is the observed pattern for all AIM1 and AIM2 scales and most AIM3 scales.

However, two AIM3 scales, Altruism and Play, have considerably lower MTMM correlations (reliabilities) than the other scales (see the column labeled AIM3 under monotrait–heteromethod in Table 8). Indeed, in the full MTMM, matrix these scales had higher monotrait–heteromethod correlations with other AIM3 scales than their MTMM correlations: Altruism positively with Legacy and Play positively with Appearance and negatively with Safety. Because this situation with Altruism and Play was not found on the AIM1 or AIM2, it seems likely that the problem arises out of the AIM3 method of measurement, rather than with these two motive constructs. Perhaps measurement of certain motives is not well served by the single-item self-endorsement method, and this should be explored in subsequent research. One other AIM3 scale with reasonable test–retest reliability (Mental) also correlates higher with another scale (Material).

The remaining MTMM matrix summary statistics for Study 7 (MTMM) are presented in Table 12. First, we consider the heterotrait–monomethod correlations of each scale with all other scales for each of the three AIM-Q methods. In the full MTMM matrix, there are 210 of these correlation coefficients for each AIM-Q method, too many to be depicted herein. Nevertheless, the summary statistics of these correlations do provide a reasonable sense of them and are found in rows labeled AIM1, AIM2, and AIM3 under heterotrait–monomethod in Table 12. The *M* values of the correlations between each of the 15 scales and all other scales within the same method are as follows: .18 ($SD = .26$) for the AIM1, .24 ($SD = .22$) for the AIM2, and .10 ($SD = .27$) for the AIM3. Campbell and Fiske (1959) suggest that these heterotrait–monomethod values should be lower than their respective MTMM and monotrait–heteromethod correlations in Table 8. The AIM-Q's heterotrait–monomethod values are considerably lower, suggesting that overall the AIM-Q scales have good discriminant validity.

Finally, we consider the heterotrait–heteromethod correlations for each pair of AIM-Q methods. Again, there are 210 of these correlations coefficients for each AIM-Q method, so the summary statistics are provided in rows labeled AIM1-AIM2, AIM1-AIM3, and AIM2-AIM3 under heterotrait–heteromethod in Table 12. The *M* values of the correlations between each of the 15 scales and all other scales for each pair of methods are as follows: .13 ($SD = .22$) for the AIM1-AIM2, .15 ($SD = .18$) for the AIM1-AIM3, and .14 ($SD = .16$) for the AIM2-AIM3. These heterotrait–heteromethod values should also be lower than their respective MTMM and monotrait–heteromethod

TABLE 12
Study 7 (MTMM) Summary Statistics: Medians, Means, Standard Deviations, Minimum, and Maximum Values of MTMM, Heterotrait–Monomethod, and Heterotrait–Heteromethod Correlation Coefficients

	Md	M	SD	*Minimum*	*Maximum*
Heterotrait–Monomethod					
AIM1	.23	.18	.26	–.40	.68
AIM1	.25	.24	.22	–.28	.72
AIM3	.10	.10	.27	–.66	.80
Heterotrait–Heteromethod					
AIM1-AIM2	.13	.13	.22	–.44	.52
AIM1-AIM3	.18	.15	.18	–.29	.46
AIM2-AIM3	.14	.14	.16	–.20	.54

correlations in Table 8, which they are. This further suggests that, overall, the AIM-Q scales have good discriminant validity.

However, a few unexpectedly large heterotrait–heteromethod correlations were observed in the MTMM matrix when the AIM3 is paired with the other two methods. For the AIM1-AIM3, the Aggression scale's heterotrait–heteromethod correlation with the Physical scale is higher than its monotrait–heteromethod (but not its MTMM) correlation with the Physical scale. Also for the AIM1-AIM3, the Altruism scale's heterotrait–heteromethod correlation with the Meaning scale is the same as its monotrait–heteromethod (but not its MTMM) correlation with the Meaning scale. For the AIM2-AIM3, the Aggression scale's heterotrait–heteromethod correlations are the same with the Mental scale and higher with the Physical scale than its monotrait–heteromethod (but not its MTMM) correlations with the Mental and Physical scales. These few problem correlations were not observed when the AIM1 and AIM2 were paired and, again, suggest problems for some constructs with the AIM3 method of measurement. With the exception of these instances, the motive constructs generally appear to have good convergent and discriminant validity as depicted in the MTMM matrix.

Study 8: Socially Desirable Responding Effects

When measuring motivational traits, it may be necessary to be aware of the socially desirable and undesirable nature of some motives. Cattell et al. (1964) were certainly concerned with this when they called for disguised, subtle measures of motivation. For example, participants may be aware that Aggression is not socially desirable, whereas Altruism is, and respond accordingly. Therefore, we explored the relationship between socially desirable responding and the motive scales as measured by the AIM1, AIM2, and AIM3.

Procedure

In study 8, different samples were used to explore relationships between social desirability and the scales of the AIM1, AIM2, and AIM3. The Internet administration procedures described in "General Procedures Used to Develop the AIM-Q" were the same, but sampling procedures differed for the AIM1, AIM2, and AIM3. The AIM1 sample ($N = 38$) was all university students, therefore there was a restricted range of ages, but there was still relatively diverse ethnic and religious identification (see Tables 2, 3, and 4 for more details). The AIM2 sample ($N = 50$) consisted of participants randomly selected from the Study 4 sample, and reflected its demographics. The AIM3 sample ($N = 66$) was obtained through the acquaintance network sampling procedures described in "General Procedures Used to Develop the AIM-Q." There was a good range of ages as well as diverse ethnic and religious identification.

A revised version of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) was used. Strahan and Gerbasi (1972) developed a Marlowe-Crowne 20-item version (M-C 20) by removing several items that contributed little to the original measure without loss of reliability. K-R 20 reliabilities for the M-C 20 ranged from .73 to .87 in four test samples. The M-C 20 was also administered online with either the AIM1, AIM2, or AIM3.

Results and Discussion

Coefficient α was calculated for the M-C 20 within each sample. The results were $\alpha = .69$ for the AIM1 sample, .78 for the AIM2 sample, and .78 for the AIM3 sample. These

coefficients are similar to those reported by Strahan and Gerbasi (1972) and suggest that the M-C 20 has adequate internal consistency in the present samples. A possible explanation for the smaller size of coefficient α for the AIM1 sample may be the restricted range inherent in an all university student sample, whereas the other two samples were more heterogeneous.

Overall, socially desirable responding does not appear to be strongly nor pervasively related to motive scale scores derived from the three AIM-Q methods. There was only one significant correlation on the AIM1 with Altruism (r = .37). The AIM1 is intended to have the most subtle items. There were two significant correlations on AIM2 with Conscience (r = .40) and Health (r = .34). AIM2 items are less subtle. There were three significant correlations on AIM3 with Appearance (r = –.24), Play (r = –.24), and Sex (r = –.30). AIM3 is the least subtle task, requiring a self-endorsement of straightforward descriptions of the motives.

GENERAL DISCUSSION

This program of studies supports the overall AIM-Q strategy of assessing individual differences in multiple evolutionary-based motives. These studies generally support the independence of the motive constructs as predicted by Bernard, Mills, et al.'s (2005) evolutionary psychology theory of motivation. Based on the MTMM results, the motive constructs appear to have good discriminant and convergent validity as measured by different AIM-Q methods. This approach is valuable because it may ultimately provide evidence of the underlying psychological constructs of the motives and, by implication, the theory on which they are based.

The present studies provide evidence of the AIM-Q motive scales' independence from each other and from social desirability, but they do not explore the AIM-Q scales' convergence with other measures or external criteria. The convergent validities of which we have spoken are those between AIM-Q motive scales in the MTMM matrix. Convergent validity must be further established by demonstrating logical, empirical relationships between the motives and external criteria. Because of the progress made in this multistudy research program, a new program of construct-validation research is now underway. The motive scales, and indirectly the constructs on which they are based, are currently being evaluated for discriminant validity with measures of intellectual functioning and for convergent validity with measures of goals, aspirations, and vocational preferences. Furthermore, sex differences in motive scores are being studied to determine if they are consistent with predictions from evolutionary theory.

One benefit of the MTMM approach has been to allow a demonstration that, with a few exceptions, the 15 motives may be reliably measured across the systematic "extraneous" influences inherent in somewhat different methods of measurement. However, this not the strongest across-methods test to which the 15 motive constructs' could be subjected. There is a degree of similarity in the methods; all three are paper and pencil tasks. A stronger multimethod test would involve more distinctly different methods such as follows: self versus other ratings, behavioral observation, and biographical information. Future research will introduce more systematic method variance inherent in distinctly different methods that should further test the validities of these constructs.

There are some differences in the homogeneity, internal consistency, and test–retest reliabilities of some of the motive scales, depending on the method of measurement. The internal consistency of two AIM1 scales (Altruism and Conscience) is lower than we would predict. The test–retest reliabilities of one AIM2 scale (Affection) and two AIM3 scales (Altruism and Play) are also lower than we would predict. In addition, the convergent validities (monotrait–heteromethod correlations) of Aggression (AIM1-AIM3 and AIM2-AIM3) and Altruism (AIM1-AIM3) are low relative to their discriminant validities (heterotrait–heteromethod correlations). Some of these unexpected results might have occurred because the measurement of certain motives may be more sensitive to context and time, as well as method. However, the other results are adequate for a new measure and the AIM1-AIM2 convergent and discriminant validities within the MTMM matrix are generally good.

Finally, socially desirable responding appears to have very little relationship with AIM-Q scores. This may appear to be a surprising finding, given the role social desirability has played in the development of many personality measures (Ben-Porath & Waller, 1992). Prominent measures, such as the MMPI (Hathaway & McKinley, 1983) and the Personality Research Form (Jackson, 1967), have included specific methods to assess the influence of socially desirable responding on scores. Also Cattell et al. (1964) had expressed particular concern about socially desirable responding when measuring motives. This was one reason we designed the three methods to vary the subtlety of items. However, the present results suggest this concern is unfounded. Indeed, Costa and McCrae (1983) have argued against including social desirability scales and corrections in personality measures, and it appears that there is little concern for socially desirable responding effects on the AIM-Q measurement methods in their current state of development.

The samples in these studies were reasonably diverse on certain criteria, and included a substantial majority of

nonstudent community adults as well as university students. However, on some conceivable criteria, such as income level, given the acquaintance network sampling method, these samples are probably not as diverse as the population at large. Another limitation of the present study is the paucity of young adults who are not university students. Sample sizes between studies also varied considerably. However, the differences in sample sizes was deliberate and logical: Large samples were used for procedures such as principal component analyses and to determine internal consistencies of the scales, whereas small samples were used for test–retest reliabilities and social desirability analyses. The latter are more demanding on participants' time and attention and should be repeated with many samples at different locations.

If the validity of the AIM-Q is supported by the new program of studies, it may offer two advantages. The first is for a growing number of evolutionary psychologists and behavioral geneticists who have sought to incorporate individual differences into their research (Buss, 2004). They have had to turn to existing self-report measures of personality (Ebstein, Benjamin, & Belmaker, 2003), none of which were designed for this use. The AIM-Q strategy may eventually offer them several new individual differences measures of motivation based specifically on evolutionary theory.

Second, if valid, the AIM-Q may also eventually offer an opportunity for applied psychologists to broaden their assessment strategies. The utility of motivational assessment is its potential for predicting future behavior. Recent attention has focused on the incremental validity of new measures for clinical uses (Haynes & Lench, 2003). When convergent validity has been satisfactorily demonstrated, the AIM-Q may offer increased incremental validity to predictions of future behavior which, up to now, have been based primarily on measures of personality and psychopathology alone. Therefore, the AIM-Q may eventually provide opportunities for measurement within applied areas such as forensic, clinical, health, and educational psychology.

REFERENCES

Ben-Porath, Y. S., & Waller, N. G. (1992). Five big issues in clinical personality assessment: A rejoinder to Costa and McCrae. *Psychological Assessment, 4*, 23-25.

Bernard, L. C., Mills, M., Swenson, L., & Walsh, R. P. (2005). An evolutionary theory of human motivation. *Genetic, Social, and General Psychology Monographs, 131*, 129-184.

Bernard, L. C., Walsh, R. P., & Mills, M. (2005). Ask once, may tell: Comparative validity of single and multiple item measurement of the Big Five personality factors. *Clinical and Consulting Psychology Journal, 2*, 40-46.

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin, 126*, 187-219.

Burisch, M. (1984a). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39*, 214-227.

Burisch, M. (1984b). You don't always get what you pay for: Measuring depression with short and simple versus long and sophisticated scales. *Journal of Research in Personality, 18*, 81-89.

Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality, 11*, 303-315.

Buss, D. M. (2004). *Evolutionary psychology*. Boston, MA: Pearson.

Buss, D. M., & Greiling, H. (1999). Adaptive individual differences. *Journal of Personality, 67*, 209-243.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.

Cattell, R. B., Horn, J. L., Sweney, A. B., & Radcliffe, J. A. (1964). *Handbook for the Motivation Analysis Test*. Champaign, IL: Institute for Personality and Ability Testing.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Cloninger, C. R. (1987). A systematic method for clinical description and classification of personality variants: A proposal. *Archives of General Psychiatry, 44*, 573-588.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

Costa, R. R., & McCrae, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882-888.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Darwin, C. (1859). *On the origin of species.* London: Murray.

Dawkins, R. (1982). *The extended phenotype.* Oxford, UK: W. H. Freeman & Co.

Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford University Press.

DeVellis, R. F. (1991). *Scale development*. Newbury Park, CA: Sage.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size, and language in humans. *Behavioural and Brain Sciences, 16*, 681-735.

Dunteman, G. H. (1989). *Principal components analysis*. Newbury Park, CA: Sage.

Ebstein, R. P., Benjamin, J., & Belmaker, R. H. (2003). Behavioral genetics, genomics, and personality. In R. Plomin, J. C. Defries, I. W. Craig, & P. McGuffin (Eds.), *Behavioral genetics in the postgenomic era* (pp. 365-388). Washington, DC: American Psychological Association.

Eid, M., & Diener, E. (2006). Introduction: The need for multimethod measurement in psychology. In M. Eid & E. Diener, *Handbook of multimethod measurement in psychology* (pp. 3-8). Washington, DC: American Psychological Association.

Eysenck, H. J. (1952). *The scientific study of personality*. London: Routledge & Kegan Paul.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

Gintis, H. (2000). Strong reciprocity in human sociality. *Journal of Theoretical Biology, 206*, 169-179.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93-104.

Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 7*, 1-52.

Hathaway, S. R., & McKinley, J. C. (1983). *The Minnesota Multiphasic Personality Inventory manual*. New York: Psychological Corporation.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*, 456-466.

Jackson, D. N. (1967). *Personality Research Form Manual*. Goshen, NY: Research Psychologists.

Kenrick, D. T., Li, N. P., & Butner, J. (2003). Dynamical evolutionary psychology: Individual decision rules and emergent social norms. *Psychological Review, 110*, 3-28.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online; report of Board of Scientific Affairs' Advisory Group on the conduct of research on the Internet. *American Psychologist, 59*, 105-117.

Mills, M. (2004, April). Evolution and motivation. In L. C. Bernard (Chair), *Integrating perspectives: Toward a new theory of human motivation.* Symposium conducted at the meeting of the Western Psychological Association, Phoenix, Arizona. [A summary is also available at the following Web site: http://www.hbes.com/intro_to_ field.htm.]

Mischel, W. (1999). *Introduction to personality* (6th ed.). Fort Worth, TX: Harcourt.

Murray, H. (1938). *Explorations in personality.* New York: Oxford University Press.

Paunonen, S. V., & Jackson, D. N. (1985a). On ad hoc personality scales: A reply to Burisch. *Journal of Research in Personality, 19*, 348-353.

Paunonen, S. V., & Jackson, D. N. (1985b). The validity of formal and informal personality assessments. *Journal of Research in Personality, 19*, 331-342.

Reese, R. J., Kieffer, K. M., & Briggs, B. K. (2002). A reliability generalization study of select measures of adult attachment style. *Educational and Psychological Measurement, 62*, 619-646.

Schmitt, M. (2006). Conceptual, theoretical, and historical foundations of multimethod assessment. In M. Eid & E. Diener (Ed.), *Handbook of multimethod measurement in psychology* (pp. 9-25). Washington, DC: American Psychological Association.

Schmitt, D. P., & Pilcher, J. J. (2004). Evaluating evidence of psychological adaptation: How do we know one when we see one? *Psychological Science, 15*, 643-649.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7*, 300-308.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191-193.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Tooby, J., & Cosmides, L. (1992). The psychological foundation of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19-136). New York: Oxford University Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35-57.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man: 1871-1971* (pp. 136-179). Chicago: Aldine.