To the Graduate Council:

I am submitting herewith a dissertation written by Paul Andrew Thoreck entitled "The Application of Conditional Reasoning to Alternate Indices of Achievement Motivation and Failure Avoidance". I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

Lawrence R. James, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

THE APPLICATION OF CONDITIONAL REASONING TO ALTERNATE INDICES OF ACHIEVEMENT MOTIVATION AND FAILURE AVOIDANCE

A Dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Paul Andrew Thoreck

August 1994

ACKNOWLEDGEMENTS

First and foremost I would like to thank God and the refining power of His Spirit whereby He enabled me to grow as a Christian throughout the dissertation process. My prayer is that this research will ultimately contribute to His greater glory. I am also thankful for the wonderful wife He has blessed me with. Marla's support and confidence and love have been (and will always be) priceless assets. And finally, I would like to thank my mentor, Dr. Larry James. Very early in our relationship it became exceedingly apparent how fortunate I was to be assigned as one of his assistants, particularly at this point in his career. I would not have wanted it any other way.

ABSTRACT

This study examined two instruments designed to test the generalizability of James' Conditional Reasoning paradigm across item formats. Consistent with James' prototypical instrument, achievement motivation and failure avoidance were the constructs of interest for both the Inferential Analogy Test (IAT) and Inferential Relations Test (IRT). Employing average test score as the performance criterion, analyses on IAT data acquired from sequential samples of Introductory Psychology students generated initial and cross-validities of .53 and .33, respectively. Composite key reliability estimation was relatively robust (.70), and indicants of construct validity were encouraging. Collectively, these results (both a priori and a posteriori) provide sufficient evidence to warrant further developmental research. More importantly, they underscore the vitality of James' conditional reasoning paradigm.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

"Conditional Reasoning" refers to the process whereby what commonly appears to be an argument (or decision) based on impartial logic and reason is subjective and susceptible to biases engendered by the latent motives of those engaged in the argument (or decision-making process). Over the past several years, Lawrence James has concentrated on examining the psychological utility of conditional reasoning, particularly as a means to better understand dispositional tendencies. The primary question facing James throughout his investigation was whether conditional reasoning could in fact be measured, and if so, would it prove to be a useful index of human temperament. Indeed, his prototypical instrument, the Conditional Reasoning Test (CRT), was designed to examine the viability of the conditional reasoning dynamic.

This manuscript centers upon a parallel initiative wherein conceptually consistent yet substantively dissimilar instruments were developed to test the generalizability of James' suppositions across item formats. Given that these instruments sufficiently embodied the principles of conditional reasoning, it was

anticipated that psychometric standards comparable to the CRT would be realized.

The Conditional Reasoning Paradigm

James' innovative approach to the measurement of personality is based upon the premise that dispositional characteristics originate and operate at the latent (i.e., unconscious) level of cognitive/affective processing. James contends that we are mostly unaware of our true motivational tendencies because at the manifest (i.e., conscious) level we are impacted by predominately pre-conscious information processing buffers (what James refers to as "justification mechanisms"), that work to enhance our self-esteem and protect our ego regardless what any given behavioral outcome may be (e.g., whether we approach or avoid an anxiety-arousing stimulus; or whether we succeed or fail on a task). These justification mechanisms may take the form of attribution errors, confirmatory biases, or any number of mental maneuvers, that nonetheless manifest themselves as relatively spontaneous and absolutely reasonable phenomena in the mind of the beholder. For example, when individuals commit the Fundamental Attribution Error (Heider, 1958; Ross, 1977), that is, they ascribe their own or another's

behavior more to dispositional rather than situational factors, it is oftentimes considered a rational inference, even when information is available that makes it apparent the situation is driving the behavior (Jones & Harris, 1967).

James has theorized that if one can identify and measure the type and/or trajectory of the justification mechanisms individuals rely upon to render opinions, report attitudes, and even more importantly, solve problems, and if one can couple these mechanisms with specific motivational, personality-based tendencies, then the potential exists to indirectly and projectively assess latent motivational tendencies/dispositions via these mechanisms.[1]

While considering various measurement strategies, James determined that since the overall dynamic is dependent upon one's perception of what is and isn't reasonable, what better way to tap into latent constructs of interest than by constructing items that collectively resemble a test of one's ability to reason. Within the context of contrasting types of reasoning, Sternberg (1984) in his Handbook of Human Intelligence states, "...in inductive reasoning, the information contained in

---

[1]See Appendix A for a representative listing of justification mechanisms.

the premises of a problem is insufficient to reach a conclusion. As a result, one can reach inductively probable conclusions, but not deductively certain ones." (p. 235) By definition, inductive problem-solving inherently provides the variability needed to differentiate among motivational tendencies. Accordingly, James concluded it would be the most functional medium for the expression of conditional reasoning.

In sum, James envisioned an instrument wherein item format would parallel the composition of items found on conventional instruments that include and/or focus exclusively upon inductive reasoning (e.g., the Graduate Management Admissions Test). However, only some of the items would "assess" inductive reasoning skills, per se. The remaining items would be designed to stimulate a response that is theoretically directed by each examinee's latent motivational tendencies (whatever they may be) via the justification mechanisms relied upon to solve the ostensibly inductive (albeit conditional) reasoning problems. In other words, the examinee would be tactically encouraged to consider the entire instrument as a test of inductive reasoning; however, the conditional reasoning items would surreptitiously stimulate the individual to reveal his/her motivational tendencies. James was convinced that if this hypothetical instrument

performed in accordance with expectations, the end result would be a theory-based and strategically derived projective technology. In pursuit of this objective, he selected two complementary constructs, achievement motivation and fear of failure, to provide investigative context.

Achievement Motivation and Failure Avoidance: Construct Explication and Conventional Measurement

Achievement motivation (AM) as a psychological construct dates back to the writings of William James (1890), wherein he discussed the importance of "achievement strivings" and the feelings of "well-being and elevation" that are derived from the attainment of self-imposed goals. J.W. Atkinson (1957, 1964, 1978) considered need for achievement in terms of the capacity for taking pride in accomplishment. His analysis of achievement behavior focused not only upon motivation to achieve, but also motivation to avoid failure. These motivational tendencies (according to Atkinson) determine whether a person will ultimately approach or avoid an achievement-oriented task. More recently, Helmreich and Spence (1978) identified preference for challenging and difficult tasks, enjoyment of working hard, and affinity

for interpersonal competition as fundamental aspects of AM; and Nicholls (1984) defined achievement behavior as that behavior in which the goal is to develop or demonstrate high ability, or to avoid demonstrating low ability.

Given that each of these perspectives (among others) offers some insight into the nature of achievement-oriented behavior, James synthesized existing theory, and in doing so identified a number of key indicators. For example, people with elevated achievement motivation have a tendency to: 1) approach demanding tasks; 2) exhibit high-intensity effort when working on a demanding task; 3) demonstrate a willingness to persist on a demanding task over time; and 4) display a positive orientation toward stressful situations and/or stimuli. Moreover, James contends that the self-efficacy of achievement-oriented individuals is typically never in doubt, and that they will most likely make an internal, effort-related attribution regardless whether they succeed or fail on a given task (see Weiner, 1972).

James (not unlike Atkinson) considers fear of failure to be a synergistically related construct, wherein manifest indicators include: 1) actively or passively avoiding responsibility for demanding tasks; 2) choosing to evade demanding tasks; 3) exhibiting ego-protective

strategies such as self-handicapping if escape/evasion from demanding tasks is not possible, and 4) displaying a negative orientation toward stress. On occasion, he has used "failure avoidance" (FA) as the operative term to underscore the expectancy that the prospect of failure is motivationally sufficient to direct the ultimate behavior of FA-oriented individuals, and perhaps more importantly the performance-related attributions these individuals make. Given its semantic salience, "failure avoidance" will be used in lieu of "fear of failure" throughout the remainder of this manuscript.

Beyond construct explication, where there is relative agreement among theorists, the measurement of achievement motivation and failure avoidance has proved to be an almost insurmountably difficult task.

Achievement motivation has been assessed projectively, most commonly via the Thematic Apperception Test (TAT, McClelland et al, 1953; Murray, 1943), and/or by way of a wide variety of self-report scales. These include instruments singularly dedicated to the construct (e.g., Mehrabian Achievement Scale, Mehrabian, 1969), as well as broad-based inventories that incorporate AM as a sub-scale (e.g., Personality Research Form, Jackson, 1967).

A relatively comprehensive review of the achievement motive and its measure was conducted by Fineman (1977), wherein he examined the convergent validity of a number of different instruments across several studies, each of which included multiple measures of the construct. Of the 78 correlations reported by Fineman, only 22 were statistically significant. The overall median correlation was .12, and the median correlation between the TAT and other projective methods was .17. Self-report questionnaires correlated .15 with the TAT, and the median correlation among questionnaire measures themselves was a mere .10.[2] Collectively, these findings suggest that something more and/or different than achievement motivation was being measured by these ostensibly AM-oriented indices.

Fineman explained the results (at least in part) by underscoring the psychometric inadequacy of many existing AM scales. Even though some of the more common self-report questionnaires fared much better than projective indices, one must move beyond the empirical to the conceptual when considering self-report measures. Nunnally (1978) has argued that self-report measures are limited to what individuals know about their attitudes,

---

[2]Conceptually similar results were reported by Weinstein in 1969.

and then are willing to relate. This speaks both to the latent and manifest levels of psychological processing. An individual's unvarnished attitude, or opinion, or self-description about something may simply not be consciously available. And if it is available, it may not be expressed due to the demand characteristics of the environment (there may be motivation to give a socially desirable response, for example).

McClelland (1985; and in collaboration with Koestner & Weinberger, 1989) has argued that self-report inventories measure not the achievement motive, but rather its "conscious counterpart", the achievement value; projective techniques are needed to tap into the achievement motive (also see Biernat, 1989). According to deCharms, Morrison, Reitman, and McClelland (1955), a key distinction between values and motives is that they predict different kinds of behaviors. Values predict "respondent behaviors", what people cognitively decide should be done, while motives predict "operant behaviors", what people will spontaneously do. Even though this line of thinking represents a commendable effort to conceptually enhance the viability of the TAT, it does not address the central issue. That is, projective options currently available are psychometrically tenuous. Nothing in the literature suggests otherwise.

The second focal construct, failure avoidance, has traditionally been measured via the Test Anxiety Questionnaire (Mandler & S.B. Sarason, 1952; S.B. Sarason & Mandler, 1952), or some derivative thereof (e.g., Test Anxiety Scale, I.G. Sarason, 1980). There are at least two key concerns when one considers existing indices of failure avoidance. First, given that they are self-report instruments, they fall victim to the same problems explicated by Nunnally. Second, there is a question of construct validity; how is it that a manifest measure of <u>test</u> anxiety broadly applies to the universe of non-test-related scenarios wherein avoidance of failure could potentially express itself? It is beyond rational argument.

Conditional Reasoning Within the Context of Achievement Motivation and Failure Avoidance

While pursuing an extensive iterative process, James cultivated a prototypical instrument that conforms explicitly to the aforementioned conditional reasoning paradigm. James' Conditional Reasoning Test (CRT) appears to be and is presented as a test of one's ability to reason by inference; however, most of the items are designed to induce examinees to differentially frame

viable response options in a manner consistent with their underlying dispositional tendencies.

Appendix B provides a conditional reasoning exemplar from the CRT. It is anticipated that individuals who are predominantly achievement-oriented will consider Alternative "c" to be the most reasonable option, because "...it identifies a dysfunction in what is, at least latently, an undesirable intervention." (James, 1994) In other words, the proposed intervention is conceptually inconsistent with AM preferences for challenging tasks and personal accountability. Conversely, FA-oriented examinees are expected to favor Alternative "a", because it "...removes the social stigma of grouping (and endorses) a social intervention that is desirable because it is both ego-protective and ego-enhancing." (ibid).

As a point of interest, this particular example characterizes a recent advance in CRT item development technology in that the remaining three alternatives are fundamentally illogical. In essence, examinees are forced to choose between two options, each of which theoretically appeals to conditional reasoning trajectories engendered by examinees' underlying dispositions. With only two potentially viable response options, the stage is set for a comparatively less confounded manifestation of differential framing/anchoring.

Notwithstanding item construction strategy, the CRT has yielded encouraging psychometric evidence in support of James' fundamental paradigm. Specifically, successive samples of undergraduate students at a major southeastern university were administered the CRT under typical classroom conditions. Using "average test grade" as the performance criterion, initial and cross-validities of .52 and .39 were obtained.

The Conditional Reasoning Test embodies only one of many techniques that can be used when presenting an inductive reasoning problem. A key alternative suggested by Sternberg is the verbal analogy (not unlike items found on the Miller Analogies Test). Conditional reasoning instruments composed of items such as these could calibrate the generalizability of James' paradigm to alternate item formats. Accordingly, two instruments were developed to complement James' CRT, the Inferential Analogy Test and the Inferential Relations Test. Consistent with the CRT, both instruments were designed to measure achievement motivation and failure avoidance. The Inferential Analogy Test (IAT) was formatted along the lines of a standard verbal analogy test, whereas the Inferential Relations Test (IRT) conceptually extended this analogy-based approach to inductive reasoning. For example:

ambition is to _____ (self-centeredness; intensity)

as risk is to _____ (venture; liability)

is a representative conditional reasoning item from the IAT, wherein examinees are instructed to select what they consider to be the most reasonable response options for both missing elements.

The item is designed to capitalize on the subconscious processes that induce achievement-oriented individuals to frame stimuli (in this case, verbal probes) differentially relative to persons whose motivational tendencies favor failure avoidance. Drawing from James' original item development technology, this expectation is based upon Atkinson's Resultant Achievement Motivation model, wherein achievement motivation and failure avoidance are presented as synergistically related constructs that anchor a shared continuum (Atkinson, 1957; 1978). More to the point, however, James provides a theoretical basis upon which to anticipate differential framing in his discussion of thematic contrasts (personal communication, 1992; 1993). In essence, James' themes represent an expansion of the construct explication process that ultimately serves as a means to strategically introduce differential framing/anchoring into the context

of inductive reasoning problems.[3] For example, the first part of the above representative item corresponds to James' "motivational salience" theme, wherein achievement-oriented individuals consider one of the principal causes of success to be intensity. Given that ambition is also a precursor to success, the "ambition-intensity" connection would be reasonable to this type of person. On the other hand, avoidance-oriented individuals are inclined to believe that their comparative lack of ability is primarily a function of limited opportunity. Thus, overt ambition is perceived as an attempt by the overprivileged to unjustly inflate themselves, which is consistent with the less than desirable characteristics of self-importance and self-centeredness.

The second part of the item exemplifies the thematic contrast between opportunity and liability (as conceived by James). Risk is framed in propitious terms by the achievement-oriented individual (e.g., as an ad*venture*), whereas in the mind of avoidance-oriented folk "one should not risk time and resources on problems which may be characterized as reckless ventures, (leading to potential) liabilities." (James, 1994)

---

[3]See Appendix C for a brief description of these themes. It should be noted, however, that James is currently engaged in a review process that will inevitably result in structural changes.

With respect to justification mechanisms per se, the presence of differential framing/anchoring has already been established. Other mechanisms, however, may be in operation as well. For example, it is expected that avoidance-oriented individuals would discount/devalue the normative schema aroused by terms such as "ambition", and perhaps even believe that (contrary to social norms) most people share this opinion (i.e., false consensus). However, given that simple verbal probes do not provide conspicuous context (as items on the CRT do), the identification of operative justification mechanisms beyond differential framing/anchoring is speculative.

Items on the Inferential Relations Test were formatted to induce respondents to select a target word that is most reasonably related to a corresponding stimulus term, and to indicate on a scale from 1 to 3 the relative strength of this relationship. To illustrate:

persistent: dedicated_____ obsessed_____

is an exemplar from the IRT. The design process for this particular item was driven by James' "connotation of stress" theme, wherein achievement-oriented individuals have a tendency to define stress in positive terms (e.g., the affect/stress one experiences when striving and

persisting to be successful is a reflection of dedication and desire). Conversely, persons who are motivated by failure avoidance are more likely to perceive stress in a predominantly negative manner. For example, given that persisting at something might induce feelings of *dis*tress, there is a predilection to define persistence in comparatively harsh terms (e.g., as an obsession).

It should be emphasized that just like the CRT, both the Inferential Analogy Test and Inferential Relations Test include ordinary inductive reasoning items as well as items designed in accordance with James' paradigm. Again, the manifest objective is to encourage respondents to consider the instruments solely as tests of inductive problem-solving. But perhaps more importantly, the items on the IAT and IRT are absolutely different in form and content from those found on the Conditional Reasoning Test, the supposition being that items designed in this manner would function as psychological probes, while maintaining the same objectives as James' conditional reasoning items. Thus, the IAT and IRT represent an independent application of conditional reasoning to alternate (and relatively elementary) indices of achievement motivation and failure avoidance, and as noted, were constructed to ascertain the boundedness of James' conditional reasoning paradigm.

CHAPTER II

METHODS

Developmental Study

The content of both the Inferential Analogy Test and Inferential Relations Test was qualitatively analyzed via verbal protocol procedures similar to those espoused by Ericsson and Simon (1984). However, an important distinction between Ericsson and Simon's protocol analysis and the process described herein is that they focused upon problems and exercises that had deductively certain solutions, whereas all of the items in the IAT and the IRT were designed to read as inductive reasoning problems.

The verbal protocol procedures were employed to achieve three distinct objectives: 1) to ensure that the instructions to examinees were uniformly interpreted, 2) to maximize the probability that both the stimulus and target terms could be found in the working vocabularies of randomly selected individuals, and 3) to ensure that the response options to the conditional reasoning items were semantically palatable (i.e., each could serve as a possible alternative).

Approximately 40 subjects were recruited to participate in the protocol analysis (i.e., 20 subjects per instrument). Subjects varied in age from 20 to 50, and included both males and females with academic credentials ranging from high school graduate to Ph.D. Subjects were employees of a major, publicly-funded institution in the northeastern United States.

Procedurally, subjects were individually asked to review the instructions to a given instrument, and to put in their own words how they interpreted the instructions. Common problem areas were modified accordingly, that is, if more than 3 subjects had any difficulty, adjustments were made[4]. Next, subjects were directed to review their respective instruments holistically, and identify words that they thought might be difficult for the "average" person to understand. Again, if more than 3 subjects identified a potential problem, an alternative was identified and re-tested. Finally, subjects were asked to focus exclusively on each conditional reasoning item, one at a time. Specifically, they were instructed to concentrate upon the target words directly across from each stimulus word, and to indicate whether or not both

---

[4]Setting the criterion at "three subjects" was somewhat arbitrary; nevertheless, it provided assurance that at least 90% of the subjects found (or ultimately did find) the instructions and items satisfactory.

target words could serve as viable response options. Subjects were not asked to distinguish correct from incorrect response options. The same correction criterion and process were applied to both instruments.

Measures

Pursuant to the protocol analysis, the first complete version of the IAT included 14 conditional reasoning items and 11 ordinary inductive reasoning items. The IRT contained 35 items, 24 of which were designed to induce conditional reasoning. Excepting that the first two items within each instrument were exclusively inductive (so as to reinforce subjects' perceptions that inductive reasoning was the sole construct of interest), items were randomly positioned, and response options within the conditional reasoning problems counterbalanced. Selected response options from the IAT were assigned a score of "1"; otherwise, they were scored "0". In like manner, conditional reasoning alternatives on the IRT were each scored as discrete items, albeit via the aforementioned three-point scale. That is, selected response options were scored "1", "2" or "3".

It was anticipated that conventional measures of achievement motivation and failure avoidance would also be of value. Given their relative popularity in the general literature, the Work and Family Orientation Questionnaire, or WOFO (Helmreich and Spence, 1978), and the Test Anxiety Scale, or TAS (Sarason, 1980) were selected. By definition, these instruments are designed to tap into diametrically distinct constructs. For example:

I find satisfaction in working as well as I can.

I usually get depressed after taking a test.

are representative items from the WOFO and TAS, respectively. A 39-item composite derived from these instruments was generated, wherein each item was scored on a standard five-point scale ranging from "strongly disagree" to "strongly agree". The WOFO/TAS composite was developed (and resultant data provided) by James' conditional reasoning research team.[5]

On the criterion side, academic (i.e., exam) performance was identified as a credible measure given that the anticipated subject pool would consist of undergraduate students, and given that "...students with a

[5]Again, it should be noted that this project parallelled James' conditional reasoning research initiative.

strong motive to achieve and who desire to (perform well) in class will (in all likelihood) devote intense effort to studying for an exam." (James, 1994) Even though this is a reasonably well-precedented strategy in achievement motivation research (Atkinson & Litwin, 1960; Spence & Helmreich, 1983), it should be noted that not everyone agrees with the underlying logic (e.g., McClelland, 1989). Nevertheless, it was determined that academic behavior aggregated over multiple occasions would enhance criterion viability (Epstein & O'Brien, 1985; Rushton, Brainerd, & Pressley, 1983), thus, "average test score" within a given course served as the principal performance measure.

As a caveat, clearly there is more to academic performance than achievement motivation, for example, intellectual ability. Therefore, no claim is being made regarding absolute prediction of the criterion. Moreover, whatever their predictive efficacy, composite keys derived from the Inferential Analogy Test and Inferential Relations Test will by definition be interpretively bounded by the fact that a specific and singular criterion was employed to establish their respective validities (James, 1973).

Subjects and Data Collection

In addition to the 40 subjects who took part in the protocol analysis process, three successive samples of 205, 175, and 415 subjects were solicited to participate in the principal phases of the study. Samples were drawn during the Fall 1992, Spring 1993, and Fall 1993 semesters, respectively. Within each sample, subjects were enrolled in an Introduction to Psychology course at a major southeastern university, and consisted primarily of Freshman and Sophomore-level men and women in their late teens to early twenties.

Data collection procedures were comparable across samples. During the latter part of a scheduled class period, students were advised by their instructor that they had an immediate opportunity to participate in a research project, and that they would qualify for extra credit if they chose to do so. They were also advised that the process involved completing an instrument designed to assess their ability to reason, and that it would only require a very limited amount of their time. In every case, all of the students in attendance agreed to participate, and under the rubric of James' conditional reasoning research initiative, access to personal demographic and academic information was authorized.

Subjects remained seated, and the two focal instruments were distributed so that every other subject received the same instrument (i.e., approximately half the subjects from each sample completed each instrument). Subjects were directed to carefully review the instructions to their test, and to raise their hands if there were any procedural questions (see Appendix D for an example of instructions to subjects). All subjects completed the task within 15 minutes.

Subjects' responses to a given instrument were manually entered into a corresponding data file. This provided an opportunity to peruse the raw data for each subject, thereby maximizing the probability that response set tendencies could be identified. Moreover, IAT subjects` responses to three particularly straightforward inductive reasoning items served as indicators of attention to task and/or reasoning ability. Consequently, more than one inappropriate response to this set of three was considered evidence of faulty cognitive processing. The data files did not include cases wherein potentially problematic response set tendencies, inattention to task and/or deficient reasoning were demonstrated.

Analytic Methodology

Hogan and Nicholson (1988) have convincingly affirmed that "...the process of test validation is hypothesis testing...(and that)...measurement-based research is formally identical with any other type of legitimate scientific inquiry." (p. 621) Accordingly, this section profiles the investigative strategy relied upon to examine the psychometric foundation and integrity of the focal instruments. First, item analytic processes are reviewed, to include specific criteria for key development. Construct validity, the "sine qua non" of validation research, is then considered, and finally cross-validation and reliability estimation.

Composite Key Development[6]

Biserial correlations were computed on subjects' responses to conditional reasoning alternatives within each item and the performance criterion (i.e., average test score). Biserial correlation was considered appropriate given that each conditional reasoning item represented an AM/FA dichotomy, underneath which existed a

---

[6]Identical item analytic and scale development procedures were applied to each of the samples.

presumedly continuous and normally distributed dimension anchored by these constructs (Atkinson, 1957, 1978; Cohen & Cohen, 1983). Alternatives that correlated with the criterion at the .10-level or better and that exhibited response rates (expressed as proportions, or "p-values") between .10 and .90 were deemed viable for inclusion in their respective composite keys. The rationale for setting alpha at .10 was based on the fact that conditional reasoning item development technology is not only new, but also projective. Thus, the potential for random fluctuations across samples was deemed higher than might be expected with conventional self-report items. Restricting p-values was founded on a similar premise, that is, items with distributive proportions outside of this range might not provide enough reliable variance to control for spurious relationships being misinterpreted as statistically and conceptually significant.

An additional criterion for composite key inclusion was whether or not the sign of the biserial correlation made theoretical sense. Specifically, each conditional reasoning alternative was designed to engender AM- or FA-specific responses, wherein a positive correlation with the criterion indicated an AM-orientation, and a negative correlation, FA. A few alternatives from each instrument, however, correlated with the criterion in a direction

opposite to expectation. In these instances, either the underlying rationale was adjusted (if indeed this was a theoretically tenable thing to do), or the item was discarded. As an added safeguard, tetrachoric correlations were computed to check for intra-item anomalies on the IAT (e.g., significant negative correlation between response options that both keyed AM or FA). A single item from the final sample proved troublesome, and thus was excluded from consideration.

Consistent with the strategy James executed in his development of CRT composite keys, results from the preceding item analysis were used to develop a unit-weighted scoring system that incorporated Atkinson's perspectives regarding resultant tendencies to approach or to avoid demanding tasks (Atkinson, 1957; 1964). As succinctly stated by James (1994), "...an estimate of the resultant tendency to approach versus to avoid demanding tasks was computed for each respondent by [a] giving the respondent a '+1' for each alternative he/she selected that keyed positively (an indicator of the motive to achieve), [b] giving the respondent a '-1' for each alternative he/she selected that keyed negatively (an indicator of the motive to avoid failure), and [c] subtracting the number of -1s from the number of +1s. The last step is identical to summing the -1s and +1s to

obtain a resultant tendency score." (p.20) These predictor-based key scores provided the necessary foundation to examine the psychometric properties of each instrument (as described in the following sub-sections).

Construct Validation

A primary concern was the construct validity of conditional reasoning itself, wherein "conditional reasoning" served as the measurement construct. According to Cook and Campbell (1979), assessing construct validity depends upon two processes: first, testing for a convergence across different measures or manipulations of the same "thing", and second, testing for a divergence between measures and manipulations of related but conceptually distinct "things". Extant from the initial sample of 205 Introductory Psychology students (i.e., the Fall 1992 sample, or F92) were three composite keys derived from three separate conditional reasoning instruments (the IAT, IRT and CRT). Also extant were conventional measures of achievement motivation and failure avoidance (specifically, the aforementioned WOFO/TAS composite was administered to F92 in order to obtain self-report estimates of these constructs).

It was anticipated that convergent validity would be evidenced among the conditional reasoning instruments given that they all shared a common focus, both in terms of the constructs used to provide context for the conditional reasoning dynamic as well as the technology employed to develop the items. Notwithstanding the fact that the self-report instruments were also designed to tap into these same contextual constructs, their item development technology was conceptually distinct. Thus, some indication of divergent, or discriminant validity was expected between the conventional instruments and the conditional reasoning instruments. The simultaneous demonstration of both convergent and discriminant validity would establish (at least the beginnings of) conditional reasoning construct validity.

Composite Key Cross-Validation

It was anticipated that the validity generalization paradigm provided by James, Ellison, Fox, and Taylor (1972) would be suitable to cross-validate composite keys (see Composite Key Development sub-section). Procedurally, composite keys developed on an independent sample of Introductory Psychology students (i.e., the Spring 1993 sample, or S93) would function as a priori

keys that could be used to score responses on a subsequent, cross-validation sample of Introductory Psychology students (i.e., the Fall 1993 sample, or F93).[7] The resultant key scores could then be correlated with the criterion from F93, thereby generating an estimate of composite key cross-validation.

Composite Key Equivalence Reliability Estimation

Drawing once again from James, Ellison, Fox, and Taylor (1972), it was anticipated that composite key reliability could be estimated by examining composite key equivalence. That is, for each instrument, composite keys from S93 *and* F93 would be used to generate key scores on the F93 data set. These key scores could then be correlated to estimate the degree to which the two keys were similar, recognizing of course that the key from S93 was a priori in F93. It is noteworthy that F93 composite keys could be developed only after composite key cross-validation was completed, given that composite key cross-validation is by definition an a priori exercise.

---

[7]An independent sample was accessed given that F92 item analytic results precipitated minor yet potentially significant changes to both instruments. No subsequent modifications were made.

Test-Retest Reliability Estimation

The inclusion of test-retest reliability estimation was intended to complement and possibly serve in lieu of composite key reliability estimation. Although test-retest is not an ideal strategy for a cognitive-oriented instrument (Anastasi, 1982), it was determined to be the most viable secondary avenue. Among the alternatives, gauging the reliability of composite keys via indices of internal consistency was not psychometrically feasible given the multi-factorial nature of achievement-motivation and failure avoidance (Cassidy & Lynn, 1989; Spence & Helmreich, 1983; James, 1994). Neither was an alternate forms strategy given that alternate forms have yet to be developed. Procedurally, it was anticipated that each instrument would be administered twice to the Fall 93 sample, with an interval of approximately four weeks separating the two administrations. Composite keys developed on this sample (i.e., F93) would be used to generate corresponding sets of key scores, which could then be correlated to estimate test-retest reliability for each instrument.

As noted, instrument-based "hypotheses" were incorporated into the preceding analytic methodology. Concomitant issues will be treated as post-hoc analyses.

CHAPTER III

RESULTS

Construct Validation

Analyses conducted under the rubric of "construct validation" were based exclusively upon data collected during the Fall 1992 (F92) semester. Sixteen (or 28%) of 56 conditional reasoning alternatives on the Inferential Analogy Test satisfied the established criteria for composite key inclusion (see Composite Key Development sub-section). Given that instrument format engendered statistical interdependency (i.e., in essence this was a forced choice exercise), eight of these keyed positive (or AM) and eight keyed negative (or FA). On the Inferential Relations Test, 14 of 48 alternatives (i.e., 29%) keyed; and even though interdependency was less of an issue (allowing that each alternative could be scored from zero to three), half keyed positive and half negative.

Regarding James' Conditional Reasoning Test, 37 (or 34%) of 110 conditional reasoning alternatives keyed successfully. His criteria for composite key inclusion, however, differed slightly in that alternatives with p-values of less than .95 or greater than .05 were eligible.

As noted, conventional measures of achievement motivation and failure avoidance were also obtained during the Fall 92 semester using abbreviated versions of the Work and Family Orientation Questionnaire and the Test Anxiety Scale, respectively. Items were scored in accordance with established protocols, and even though internal consistency as a measure of reliability was considered inappropriate for the IAT and IRT, it proved functional for the more traditional instruments (i.e., both the WOFO and TAS exhibited coefficient alphas of .82)[8] Descriptive statistics corresponding to each of the predictors and the performance criterion are provided in Table 1.

As anticipated, item composites from the focal instruments correlated comparatively well with the Conditional Reasoning Test (vis a vis the WOFO and the TAS). Given that the simultaneous demonstration of convergence across different measures of the same "thing" and divergence between measures of related but conceptually distinct "things" is prima facie evidence for construct validity, "conditional reasoning" seems to have merit as a construct in and of itself. Moreover, the conditional reasoning composite keys exhibited superior

---

[8]Across samples, coefficient alpha did not exceed .40 for either the Inferential Analogy Test or the Inferential Relations Test.

predictive efficacy on the criterion relative to the conventional instruments. Table 2 documents relevant inter-correlations among the predictors, as well as predictor-criterion correlations.

Interestingly, the magnitude of the correlations between the conditional reasoning instruments suggested that they might be differentially measuring previously explicated and/or undefined aspects of the contextual constructs. In an effort to ascertain the existence of complementary variance, two separate analyses were conducted wherein average test score was regressed on F92 CRT and IAT key scores, and then on CRT and IRT key scores.

Table 1

F92 Means, Standard Deviations, and Sample Sizes

	Mean	SD	Sample Size
IAT Composite Key	-1.27	2.51	104
IRT Composite Key	-1.26	4.90	101
CRT Composite Key	2.21	3.69	377
WOFO (19 items)	68.64	8.59	315
TAS (20 items)	62.01	10.32	306
Average Test Score (ATS)	39.21	3.99	422

Table 2

F92 Correlation Matrix

	IAT	IRT	CRT	WOFO	TAS	ATS
IAT	1.00	_	.32** (N=89)	-.04 (N=82)	-.19 (N=81)	.52*** (N=98)
IRT		1.00	.47*** (N=88)	.01 (N=75)	-.02 (N=74)	.61*** (N=94)
CRT			1.00	.07 (N=283)	-.26*** (N=274)	.52*** (N=358)
WOFO				1.00	-.12* (N=304)	.11 (N=309)
TAS					1.00	-.24*** (N=309)
ATS						1.00

* p<.05
** p<.01
*** p<.001

Findings from the first analysis indicated that regardless which predictor was first considered in the regression sequence, both the CRT and IAT incrementally contributed approximately 11% to the total variance accounted for (i.e., 34%). The results from the second analysis were strikingly dissimilar, that is, the CRT incrementally contributed only 1% to the total variance beyond that apportioned to the IRT (i.e., 41%); and the

[9]Although comparatively low, it is not surprising that the Test Anxiety Scale was significantly predictive of the criterion (given their conceptual correspondence).

IRT, 28%. Acknowledging that these findings were subject to single sample anomalies, it is interesting to note that both the IAT/CRT and CRT/IAT regression sequences substantially improved predictive efficacy, thereby suggesting the prospect of using both instruments in testing situations to enhance overall prediction. However, extensive cross-validation research would be necessary so as to confirm this dynamic.

Hierarchical regression analyses were also executed to ascertain whether the IAT and/or the IRT accounted for unique variance beyond that which could be apportioned to a variety of predictors, including critical thinking ability (as measured by ACT). Table 3 provides descriptive and correlational information pertaining to subjects' ACT scores. Tables 4 and 5 document the hierarchical regression results.

Notwithstanding the fact that four viable predictors preceded the focal instruments in the hierarchy, both the IAT and IRT substantially improved prediction on the criterion. The magnitude of these results, however, gives rise to an even more fundamental question, specifically, "Will this predictive potency maintain itself across samples, particularly when cross-validation composite keys are used to generate key scores?" Pertinent findings are presented in Tables 9 and 10.

Table 3

F92 ACT Descriptive Statistics and Zero-Order Correlations

ACT Descriptive Statistics	
Mean	22.88
Standard Deviation	3.45
Sample size	413
ACT Zero-Order Correlations	
IAT Key Scores	.08 (N=97)
IRT Key Scores	.46*** (N=91)
CRT Key Scores	.44*** (N=348)
WOFO	.07 (N=298)
TAS	-.26*** (N=289)
Average Test Score	.49*** (N=398)

*** p<.001

Table 4

F92 IAT Hierarchical Regression

	R Square Change (sr^2)	Cumulative R Square
ACT		.18***
CRT	.07*	.25***
WOFO	.02	.27***
TAS	.03	.30***
IAT	.09**	.39***

* p<.05
** p<.01
*** p<.001
N=67

Table 5

F92 IRT Hierarchical Regression

	R Square Change (sr^2)	Cumulative R Square
ACT		.28***
CRT	.04*	.32***
WOFO	.00	.32***
TAS	.00	.32***
IRT	.20***	.52***

* p<.05
*** p<.001
N=65

Composite Key Cross-Validation

Composite keys used for cross-validation purposes in Fall 1993 were developed on data accessed during the Spring 1993 (S93) semester. For this sample, 43% of the conditional reasoning alternatives on the Inferential Analogy Test keyed successfully (12 positive and 12 negative). On the Inferential Relations Test, 10 of 48 (i.e., 21%) satisfied the established criteria (7 positive, and 3 negative). Means, standard deviations, and sample sizes for each of these composite keys, subjects' ACT scores, and the criterion are provided in Table 6. Correlations among these variables are documented in Table 7.

Table 6

S93 Means, Standard Deviations, and Sample Sizes

	Mean	SD	Sample Size
IAT Composite Key	1.81	3.28	97
IRT Composite Key	4.65	4.42	78
ACT Score	22.33	3.92	194
Average Test Score	37.24	4.46	265

Table 7

S93 Correlation Matrix

	IAT	IRT	ACT	ATS
IAT	1.00	-	.31** (N=68)	.58*** (N=94)
IRT		1.00	.46*** (N=51)	.43*** (N=74)
ACT			1.00	.35*** (N=176)
ATS				1.00

** p<.01
*** p<.001

The correlational results clearly indicate that the S93 initial validity estimate for the IAT was comparable to that obtained on the F92 sample, whereas it noticeably declined for the IRT. Moreover, relative to traditional indices of achievement motivation and failure avoidance (Fineman, 1977; Weinstein, 1969), the IAT exhibited robust predictive efficacy, particularly in the cross-validation arena. Composite key cross-validities computed on Fall 1993 (F93) data are provided in Table 8.

In the preceding section it was suggested that hierarchical analyses could provide a more rigorous test of predictive efficacy (wherein IAT and IRT key scores were generated via cross-validation composite keys). Accordingly, (for each instrument) the criterion was

regressed on F93 subjects' ACT scores and their corresponding cross-validation key scores.[10] These results are documented in Tables 9 and 10.

Table 8

S93 --> F93 Composite Key Cross-Validities

	Average Test Score
IAT Composite Key	.33*** (N=190)
IRT Composite Key	.15* (N=191)

* p<.05
*** p<.001

Table 9

F93 IAT Hierarchical Regression

	R Square Change (sr^2)	Cumulative R Square
ACT		.30***
IAT	.05***	.35***

*** p<.001
N=181

[10]Given that ACT proved to be substantively and significantly predictive of the criterion in the Fall 92 sample, no other variables (beyond IAT and IRT) were included in these analyses.

Table 10

F93 IRT Hierarchical Regression

	R Square Change (sr^2)	Cumulative R Square
ACT		.29***
IRT	.00	.29***

*** p<.001
N=180

It is striking that even under stringent cross-validation conditions, robust predictive efficacy on the Inferential Analogy Test was attained (re: Table 9). This suggests that academic selection practices could be productively augmented with conditional reasoning-based instruments. However, once again the Inferential Relations Test exhibited questionable stability, an issue that will be more thoroughly addressed in the next chapter.

Post hoc examination of the composite keys identified a pair of alternatives from the Inferential Analogy Test that keyed on both the initial validation and cross-validation samples, yet in different directions. Just as curious was the fact that across samples, the response rates (or p-values) for these alternatives were almost identical. Unfortunately, this phenomenon was impossible to detect until after the analyses were run (given that

cross-validation is by definition an a priori exercise). Nevertheless, a priori removal of these problematic alternatives would have improved IAT cross-validity from .33 to .37. Although cross-validation of this revised composite key on an independent sample would at some point be advisable, the stability of this "corrected" validity coefficient can be partially examined via postdictive cross-validation of the Fall 93 composite key on the Spring 93 sample. Accordingly, the analysis was executed and a cross-validity of .43 was obtained.

Composite Key Equivalence and Test-Retest Reliability Estimation

Data acquired during the Spring 1993 and Fall 1993 semesters were used to estimate instrument reliability. Regarding F93, 24 of 56 conditional reasoning alternatives from the IAT keyed on the performance criterion (i.e., 43%). On the IRT, 13 of 48 (or 27%) were eligible for composite key inclusion (7 positive and 6 negative). Descriptive statistics for F93 are listed in Table 11, and Table 12 documents inter-correlations among these variables.

Table 11

F93 Means, Standard Deviations, and Sample Sizes

	Mean	SD	Sample Size
IAT Composite Key	3.14	3.07	207
IRT Composite Key	1.03	5.34	208
ACT Score	22.56	3.62	437
Average Test Score	37.73	4.32	432

Table 12

F93 Correlation Matrix

	IAT	IRT	ACT	ATS
IAT	1.00	-	.33*** (N=185)	.46*** (N=189)
IRT		1.00	.43*** (N=188)	.34*** (N=191)
ACT			1.00	.52*** (N=411)
ATS				1.00

*** p<.001

Composite key equivalence reliability estimates for both instruments offered modest evidence in support of cross-sample consistency. On the other hand, only the IRT demonstrated adequate test-retest reliability (see Table 13). Not surprisingly, a priori removal of the alternative pair noted in the preceding section would have improved the IAT composite key reliability estimate from .70 to a more acceptable .83.

Table 13

Composite Key Equivalence and Test-Retest Reliabilities

	Composite Key Equivalence	Test-Retest
IAT Composite Key	.70*** (N=206)	.25** (N=123)
IRT Composite Key	.59*** (N=208)	.67*** (N=146)

** p<.01
*** p<.001

CHAPTER IV

DISCUSSION

At the outset, it was proposed that psychometric standards comparable to those evidenced by James' Conditional Reasoning Test might be attainable with alternate instruments if indeed they sufficiently embodied the principles of conditional reasoning. Consequently, it was affirmed that initial and cross-validities on the CRT were .52 and .39, whereas on the Inferential Analogy Test .53 and .33 were obtained. Even though James employed a more rigorous criterion to establish his cross-validation composite key (i.e., alternatives had to key on two independent samples), these statistics are remarkably consistent, and in and of themselves reinforce the construct validity of conditional reasoning. However, this same consistency also suggests possible limits on the efficacy of conditional reasoning as an assessment mechanism (given the current state of item development technology). Nevertheless, relative to existing standards, wherein initial predictive validities accounting for 10% of the variance are considered superior, conditional reasoning-based instruments offer substantive promise.

It must be emphasized; however, that conditional reasoning research is at best developmentally-oriented, and for some time will be subject to an evolving theoretical foundation. James himself asserts that we are only in the beginning stages of cultivating a new measurement system. Therefore, any and all inferences must be tempered by openly acknowledging these constraints.

This research is also limited by the fact that samples were accessed primarily because they were so accessible. Accordingly, generalizing the results beyond this class of subject (or beyond the criterion, for that matter) would be questionable. Moreover, given that items on both instruments presented a forced-choice exercise, statistical interdependency was unavoidably engendered. Nevertheless, these types of problems were considered secondary to the overall objective of exploring the feasibility of alternate item formats (particularly given that James' has not even addressed these limitations). There are plans; however, to study conditional reasoning in non-academic environments using non-academic criteria; and over the next few months statistical concerns, including intra-item interdependency, will become the focus of intense examination.

More immediate to this study is the question of differential psychometric integrity across instruments. Why did the IAT seem to perform better than the IRT, particularly across samples?[11] There are several possibilities, some or all of which may provide a solution. Perhaps the differences can be attributed to the probability that the IRT was comparatively less inductive. Or possibly the results were influenced by the fact that the IAT afforded an opportunity to exclude subjects who exhibited inattentiveness and/or deficient reasoning (see Subjects and Data Collection section, p. 23). Or, perhaps subjects who completed the IAT benefitted from some sort of priming phenomenon (Feldman & Lynch, 1988), given that each item had two distinct elements. Research designed to examine these hypotheses could in fact contribute to our overall understanding of conditional reasoning.

A somewhat related matter is the fact that the IAT performed so poorly on test-retest. This could be construed as an indicator that something is fundamentally awry. However, it could also be that the dismal test-retest reliability estimate on the Inferential Analogy Test was more a function of technique rather than psychometric deficiency. Comparatively speaking, the IAT

---

[11]See Tables 8-10 and 13.

presented a more demanding cognitive processing requirement than did the Inferential Relations Test, simply by virtue of the fact that each item had two distinct parts that somehow were supposed to be interrelated. However, there was no logical relationship, at least none that had been intentionally introduced. Accordingly, the retest experience may have put an even heavier cognitive load on subjects to look for logical interrelationships so as to better remember how they had responded four weeks prior. Searching for logical relationships where none existed could very well have resulted in confusion and diminished test reliability, whereas the first time they took the test it might have been more of a visceral experience, which indeed was the original objective. The fact that adequate composite key equivalence was demonstrated would seem to support this interpretation. Moreover, these results (at least in part) could account for the IAT's superior cross-validation performance relative to the IRT, that is, the Inferential Analogy Test (as suggested above) may indeed have been more of a "conditional reasoning-based instrument" than the Inferential Relations Test.

Another immediate concern is the identification of commonalities across alternatives that keyed on the criterion. What set these alternatives apart from the

conditional reasoning alternatives that did not key? Conceptually, they do not appear to be theme-specific; however, if some distinctive aspect does exist, it could be used to facilitate the effective design of future items. Perhaps the issue is one of criterion specificity, that is, there may be something peculiar about the conceptual relationship between the criterion and the key alternatives. Once again, additional research using alternate criteria is needed.

Aside from basic research, the potential applications of conditional reasoning item development technology extend well beyond the theoretical. Executive appraisal, an area with a tenuous history, could benefit from an instrument that is genuinely capable of identifying personality predispositions, particularly if certain traits have been associated with successful job performance. As has already been suggested, assessing one's motivation to complete graduate school, especially doctoral-level programs, could provide selection committees with an indicator that is potentially more salient and cost-effective than traditional predictors. This same reasoning could easily be extended to measuring a high school student's motivation to complete undergraduate training. Moreover, it is anticipated that complementary psychological constructs (e.g., aggression,

reliability, and power) will be integrated into this same conditional reasoning paradigm. The long-term objective will be to create an omnibus personality instrument designed for broad-based application that is relatively free from the vicissitudes of existing techniques.

Clearly, much needs to be done to scientifically establish conditional reasoning as a viable approach to the assessment of human motivation and temperament. This manuscript has simply documented an initial and independent attempt to examine the dynamic beyond the formative context James used to test his original theory. Even though conditional reasoning research is developmentally-oriented, the "early returns" are promising, and perhaps indicative of a soon to be realized paradigm shift in projective item development technology.

LIST OF REFERENCES

LIST OF REFERENCES

Anastasi, A. (1982). Psychological testing. New York: MacMillan.

Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. Psychological Review, 64, 359-372.

Atkinson, J.W. (1964). An introduction to motivation. Princeton: Van Nostrand.

Atkinson, J.W. (1978). The mainsprings of achievement-oriented activity. In J.W. Atkinson & J.O. Raynor (Eds.), Personality, motivation, and achievement (pp. 11-39). Washington, DC: Hemisphere Publishing Corporation.

Atkinson, J.W., & Litwin, G.H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. Journal of Abnormal and Social Psychology, 60, 52-63.

Biernat, M. (1989). Motives and values to achieve: Different constructs with different effects. Journal of Personality, 57, 69-95.

Cassidy, T., & Lynn, R. (1989). A multifactorial approach to achievement motivation: The development of a comprehensive measure. Journal of Occupational Psychology, 62, 301-312.

Cohen, J., & Cohen, P. (1983). Applied regression/ correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston, MA: Houghton Mifflin Company.

deCharms, R., Morrison, H.W., Reitman, W.R., & McClelland, D.C. (1955). Behavioral correlates of directly and indirectly measured achievement motivation. In D.C. McClelland (Ed.) Studies in motivation (pp. 414-423). New York: Appleton-Century-Crofts.

Epstein, S., & O'Brien, E.J. (1985). The person-situation debate in historical and current perspective. Psychological Bulletin, 98, 513-537.

Ericsson, K.A., & Simon H.A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Feldman, J.M., & Lynch, J.G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. Journal of Applied Psychology, 73, 421-435.

Fineman, S. (1977). The achievement motive construct and its measurement: Where are we now? British Journal of Psychology, 68, 1-22.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Helmreich, R.L., & Spence, J.T. (1978). The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. JSAS Catalog of Selected Documents in Psychology, 8, 35.

Hogan, R., & Nicholson, R.A. (1988). The meaning of personality test scores. American Psychologist, 43, 621-626.

Jackson, D.N. (1967). Personality Research Form Manual. New York: Research Psychologists Press.

James, L.R. (1973). Criterion models and construct validity for criteria. Psychological Bulletin, 80, 75-83.

James, L.R. (1994). Conditional reasoning: A new approach to the measurement of personality. Manuscript submitted for publication.

James, L.R., Ellison, R.L., Fox, D.G., & Taylor, C.W. (1972). Reliability estimated by scoring key equivalence across samples. Journal of Applied Psychology, 56, 500-505.

James, W. (1890). Principles of psychology. New York: Holt.

Jones, E.E., & Harris V.A. (1967). The attribution of attitudes. Journal of Experimental Social Psychology, 3, 1-24.

Mandler, G., & Sarason, S.B. (1952). A study of anxiety and learning. Journal of Abnormal and Social Psychology, 47, 166-173.

McClelland, D.C. (1985). How motives, skills, and values determine what people do. American Psychologist, 40, 812-823.

McClelland, D.C. (1989). Human motivation. New York: Cambridge University Press.

McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. (1953). The achievement motive. New York: Appleton-Century-Crofts.

McClelland, D.C., Koestner, R., & Weinberger J. (1989). How do self-attributed and implicit motives differ? Psychological Review, 96, 690-702.

Mehrabian, A. (1969). Measures of achieving tendency. Educational and Psychological Measurement, 29, 445-451.

Murray, H.A. (1943). Thematic apperception test manual. Cambridge, MA: Harvard University Press.

Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological Review, 91, 328-346.

Nunnally, J.C. (1978). Psychometric theory. New York: McGraw-Hill.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.) Advances in experimental social psychology: Volume 10. New York: Academic Press.

Rushton, J.P., Brainerd, C.J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. Psychological Bulletin, 94, 18-38.

Sarason, I.G. (1980). Test anxiety: Theory, research, and applications. Boston: L. Earlbaum Associates.

Sarason, S.B., & Mandler, G. (1952). Some correlates of test anxiety. Journal of Abnormal and Social Psychology, 47, 810-817.

Selye, H. (1956). The stress of life. New York: McGraw-Hill.

Spence, J.T., & Helmreich, R.L. (1983). Achievement-related motives and behaviors. In J.T. Spence (Ed.), Achievement and achievement motives: Psychological and sociological approaches (pp. 7-74). San Francisco: W.H. Freeman.

Sternberg, R.J. (1984). Reasoning, problem solving, and intelligence. In R.J. Sternberg (Ed.), Handbook of human intelligence (pp. 225-254). Cambridge University Press.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R.M. (1972). Perceiving the causes of success and failure. In E. Jones, D. Kanouse, H. Kelly, R. Nisbett, S. Valins, & B. Weiner (Eds.) Attribution: Perceiving the causes of behavior (pp. 95-120). Morristown, NJ: General Learning Press.

Weinstein, M.S. (1969). Achievement motivation and risk preference. Journal of Personality and Social Psychology, 13, 153-172.

APPENDICES

Appendix A

Justification Mechanism Exemplars
Anchoring
Contrast/Similarity Errors
Discounting/Devaluation
Differential Encoding
Downward Comparisons
False Consensus
Implicit Theories
Projection
Rationalization
Self-Enhancement
Stereotyping

(Source: James, 1994)

Appendix B

Exemplar from James' Conditional Reasoning Test

Students are often separated by academic ability in today's elementary schools. For example, fourth-grade students might be split into higher, middle, and lower ability classes. The main reason for separating students into ability groups is to design a specific plan of study for each ability group that meets the needs of the students in that group. Recently, however, grouping by ability has become a source of controversy because students in the lower-ability groups have been shown to have lower self-esteem and motivation to perform academically than students in the higher-ability groups. It has been recommended that grouping by ability be stopped and that every class include students from all ability levels.

Which one of the following is most likely to occur if this recommendation is implemented?

a. If students from all ability levels are meshed together, then the higher-ability students could serve as role models for the lower-ability students and many of the lower-ability students would try harder.

b. The United States would begin to catch up with both Japan and Europe in regard to scores on academic tests because lower-ability students would begin to score more highly.

c. Many advanced materials now given in the higher-ability classes would not be given to classes comprised of students from all ability levels.

d. Putting lower-ability students in classes with higher-ability students would require more classrooms because more classes would have to be taught.

e. Higher-ability students would be more likely to register for nonacademic courses such as shop and art.

(Source: James, 1994)

Appendix C

Synopsis of James' Themes

1. Personal Control and Responsibility

AMs seek responsibility and self-determination on demanding tasks because personal control engenders the attribution (by the self and others) of having been responsible for success. Differential (i.e., competitive) reward systems are desired because such systems recognize each individual's contribution. Moreover, AMs expect to be able to control aversive stimuli. They believe failure on a complex task is a temporary condition that can be rectified by changes in strategy or by increases in intensity and/or persistence. FAs are more inclined to avoid responsibility on demanding tasks, and prefer reward systems that minimize individual differences (e.g., structured environments that de-emphasize task difficulty/challenge). FAs believe that intensity, persistence, and sacrifice will not help on demanding tasks because the external, stable causes of failure, or anticipated failure, are uncontrollable.

2. Opportunity versus Liability

AMs believe that demanding and valent tasks furnish challenges and opportunities for self-enhancement, and are attracted to such tasks. Competence is amplified by working on increasingly difficult objectives. FAs anticipate failure on these same tasks and thus believe them to be threats. They desire to escape from such situations and to avoid similar situations in the future. Personal security is a key objective.

3. Connotation of Stress

Stress has a positive connotation for AMs, and is associated with intensity, persistence, success, and satisfaction. Stress is evidence that someone is trying to be competent. Stress has a negative connotation for FAs, and is associated with compulsiveness, overload, immaturity, and/or unhappiness. This distinction is conceptually comparable to Hans Selye's notion of

"eustress" and "distress".

4. Motivational Salience

AMs perceive the probability of completing a demanding, achievement-oriented task as primarily a function of persistence (i.e., dedication). Implicit is the belief that they possess sufficient ability to be successful should they choose to devote the effort. Moreover, successful people are used as role models (i.e., AMs make upward comparisons); thus, self-esteem and self-confidence are implied. FAs make downward comparisons that serve to protect and/or enhance their ego (a conceptual consistency with AMs). Persistence is devalued as being obsessive, and FAs believe that a principal cause of failure on demanding tasks is not being naturally or opportunistically imbued with prerequisite abilities and skills.

5. Compensation

One means FAs have of coping with their fear that they will fail on demanding tasks is to compensate or offset expected failure by shifting to less demanding tasks. This allows FAs to replace the expectations of demanding tasks with the diminished expectations of less exacting tasks. FAs ascribe positive attributes to the less demanding tasks (and to themselves) in order to make the tasks (and themselves) more socially acceptable, if not commendable. For example, compassion, empathy, and concern are important task-related characteristics in the FA-oriented mind. Conversely, AMs tend to frame roles and tasks with diminished expectations as dull, boring, and routine. Although traits such as empathy and compassion are socially valued, AMs discern these as emotional ploys when used to avoid demanding tasks (and embrace less demanding tasks).

Appendix D

Inferential Analogy Test Instructions to Subjects

Each item below has two missing elements. Your task is to circle what you consider to be the most reasonable response options for both missing elements within each item. Here is an example:

Book is to _____ (cover; library)
as house is to _____ (neighborhood; home)

The most reasonable response options for this item would be "library" and "neighborhood". Even though a book has a cover and a house is oftentimes considered a home, the most striking relationship between "book" and "house" is the fact that they can both be part of larger groupings of books and houses (i.e., a book can belong to a library and a house can belong to a neighborhood).

For each item below, circle what you consider to be the most reasonable response option for each missing element. Please answer all of the items. There is no penalty for guessing. Also, please print your name and student identification number (i.e., typically your social security number) in the spaces provided.

NAME_____

STUDENT IDENTIFICATION NUMBER_____

VITA

Paul Andrew Thoreck was born in Syracuse, New York, on January 13, 1956. He attended primary and secondary schools in the Syracuse area, and subsequently served four years in the United States Army. In 1981, he received a Bachelor of Science degree in Psychology from the University of Texas at El Paso, and in 1983 was awarded a Master of Arts degree in General/Experimental Psychology from the same institution. He spent the next several years working in both the public and private sectors, and in 1990 entered the University of Tennessee's Industrial/ Organizational Psychology doctoral program. Upon completion of his degree, he plans on relocating to the St. Augustine, Florida area with his family.