This is to certify that the

dissertation entitled

On the Meaning and Measurement of Test Appropriateness

presented by

Jose Manuel Cortina

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Psychology

Neal Schmitt
Major professor

Date 7/12/94

ON THE MEANING AND MEASUREMENT OF TEST APPROPRIATENESS

By

Jose Manuel Cortina

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1994

ABSTRACT

ON THE MEANING AND MEASUREMENT OF TEST APPROPRIATENESS

By

Jose Manuel Cortina

Recent research has shown that a test can be psychometrically valid and yet be inappropriate for certain individuals such that the test scores for these individuals cannot be interpreted as accurately indicating their standing on the construct of interest. This body of research, however, has been largely statistical in nature, with a focus on indices of appropriateness such as the l_z index developed by Drasgow, Levine, and their colleagues. The purpose of the present paper was to examine appropriateness as a construct, and develop and partially test a model of its determinants based on literature from educational, social, personality, and quantitative psychology. Specifically, the effects of item characteristics, math anxiety, test anxiety, carelessness, and conscientiousness on statistical knowledge test scores and the l_z index of test appropriateness were examined in a sample of 165 undergraduate statistics students. The results showed that item characteristics, math anxiety, carelessness, and the item characteristic by conscientiousness interaction were significantly related to knowledge test scores while none of the hypothesized predictors of l_z were significantly related to it. Implications for appropriateness and testing are discussed.

ACKNOWLEDGEMENTS

I've been waiting a long time to write my acknowledgements, if for no other reason than because it would suggest that I have something to acknowledge.

And I do.

It appears that I have finally managed, over the shrill objections of the administration, to annex a series of degrees up to and including Doctor of Philosopy from Michigan State University. How and when this happened, I have no idea, but there are many people who deserve recognition: accessories after the fact, if you will, and I think you will.

Thanks Kim for things too numerable to mention, but specifically for reminding me that the writing of my Results section was not an insurmountable task. You were right. Thanks Ron for Sega and Scotch and other friendship-related esoterica. Thanks Stephen and Dale for not letting trifles like work get in the way of beer and golf. Thanks Mick for reminding me not to take life too terribly seriously. After all, there's a 50-50 chance that I won't survive the day anyway. Thanks Sher for friendship unconditional upon the level of my idiocy. Thanks Stan and Jean (Jan and Stean?) for Biscotti and access to D's closet. Thanks Tim and Mo for French wine, artichoke dip, and kindness. Thanks John for playing up my golf skills in the letter of rec. Thanks Mike, Steve, Dan, and Kevin for sitting through all of those defense meetings without laughing, at least not in my presence. Thanks Mary for not getting too upset when I wallpapered your desk. Thanks Suzy, Greta, and Cheryl for always making life easier. Also, special appearances

by Jeff S., Sandy L., JoAnn S., Whit, El, Jen, Dennis, J.T., Smitty, Rob, Burker, Dan. W., Hatrack, Sandy T., Kara S., Rick D., Gordon W., and a player to be named later.

Thanks to my family for love and the absence of worry.

And thanks most of all to Neal. There aren't many people in this field or any other who could have put up with me through a thesis, comps, a dissertation, 818, Personnel Selection, 818 again, and everything in between (e.g., undocumented book and journal theft, unannounced interruptions, incessant questions, an overloaded computer account, typos, etc.). You are not only the finest mentor I have ever seen or heard about, you are, as far as I can tell, the best possible mentor, especially for a person with my particular eccentricities. My career goal, though unattainable, is to be the Academic that you are.

JMC 6/94

P.S. Special thanks to Ronald Wilson Reagan, who never let me down.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

The topic of this paper is test appropriateness. In general terms, a test is appropriate for a given individual to the extent that it measures the construct or constructs that it is supposed to measure and nothing else. Although there is a wealth of research on the determinants of test scores (e.g., test anxiety, response biases, motivation, item wording, etc.), there is a relative paucity of research on the determinants of test appropriateness. The goal of this paper is to develop and partially test a model of test appropriateness based on literature from I/O psychology, educational psychology, education, and quantitative psychology. Although some of the issues that are discussed could be applicable to tests of any kind, I focus only on multiple choice tests. Nevertheless, I make an attempt to include a wide range of test content, both maximum performance measures (i.e., tests composed of items with possible responses that are either absolutely correct or absolutely incorrect such as mathematics knowledge, reading ability, paragraph comprehension, spatial relations, etc.) and typical performance measures (i.e., tests composed of items with responses that are not necessarily right or wrong, such as personality inventories, interest inventories, etc. It should be noted, however, that typical performance tests can have right and wrong answers in a sense when they are used for selection purposes). At the outset, one note of clarification is in order. Although discussions of appropriateness are perhaps best directed at the individual item (since this is where our attempts to measure constructs with tests begin), the terms "test

inappropriateness" and "item inappropriateness" are often used interchangeably in this paper. The reason for this is simply that a test is nothing more than a set of items. To the extent that those items are inappropriate, the test composed of them is obviously inappropriate.

I begin with an overview of appropriateness as it has been studied, and follow with a review of the relevant literature, the purpose of which is to develop a model of test appropriateness.

Test appropriateness as it has been studied

A test is inappropriate to the extent that it measures constructs other than the construct of interest. A test may be inappropriate, however, only for certain respondents. For example, consider a psychometrically sound paper and pencil test of English comprehension. For most respondents, this test will yield scores that accurately reflect the English comprehension of the respondents. In other words, the test would be appropriate for these respondents. Now consider the performance of a visually impaired individual on this paper and pencil test. This respondent would almost certainly score very poorly on this test, not because of a lack of English comprehension, but because this respondent can not cope with the format of the test. For this reason, this test would be inappropriate as a measure of English comprehension for this individual.

Research on appropriateness has focussed primarily on the development of techniques that identify respondents for whom a given test is inappropriate. Specifically, these techniques involve the identification of response patterns that are aberrant and, therefore, suggest inappropriateness. One way of describing the logic of these indices is in terms of Guttman vectors. A Guttman vector is simply a vector of zeroes and ones in

which all of the ones precede all of the zeroes. If a respondent were to respond to items in a way that matched perfectly with the difficulties of the items (and if there were no possibility of guessing), then the responses of that respondent, when ordered in terms of item difficulty, would form a perfect Guttman vector. The idea is that the respondent answers all items at or below a certain difficulty level correctly. At some point, however, the difficulty becomes too great for that respondent, and all items above that level of difficulty are answered incorrectly. If this were the case, then this respondent should receive a perfect score on an appropriateness index. If, however, some of the items measured constructs other than the construct of interest, then the responses of a given individual might depart from a Guttman vector, and the responses of this person would then be "flagged" by an index of inappropriateness. Consider again the example of a visually impaired individual taking a test of English comprehension, except that now there are some paper and pencil items and some items that are asked and answered in an interview format. For those respondents without impaired vision, we would expect little difference between the written questions and the interview questions. As a result, we could order all of the dichotomously scored item responses for these respondents in terms of their difficulty values and expect them to form something resembling a Guttman vector such as this

1111111100000000

The visually impaired individual, however, would almost certainly do much better on the interview items regardless of their group-determined difficulty values.

This individual, therefore, would have an "aberrant" pattern of responses such as the following

001000110010110

Almost all of the 1's for this individual could be expected to represent interview items. Since these interview items as a group should have levels of difficulty similar to those of the paper and pencil items, there should be items of both types at all levels of group-determined difficulty. For the visually impaired individual, however, the most prominent source of "difficulty" is whether or not the items must be seen to be answered. In other words, there is a source of item difficulty for the visually impaired respondent that does not apply to the group on which the item difficulties were determined.

While this is a useful example for explication, it is an exaggeration. More realistic examples would be mathematical word problems (or word problems of any kind) given to people who cannot read well, items with "culture-loaded" content given to someone unfamiliar with the culture, and any knowledge, ability, or personality test given to someone with extreme test anxiety.

Many indices have been developed that identify aberrant response patterns, such as Sato's Caution Index (Sato, 1975), the Dependability Index (Kane & Brennan, 1980), and the l_z index (Drasgow, Levine, and Williams, 1985). The l_z index, however, has received the most recent attention and is, therefore, the appropriateness measure to be used in the present study. Although the specifics of the index are described later in the paper, the general purpose of l_z is to assess the extent to which the responses of a given respondent conform to the three-parameter Item Response Theory (IRT) model. This is analogous to the Guttman-based indices such as those of Sato (1975) and Kane &

Brennan (1980), except that the l_z index assesses the congruence between the responses of an individual and the item parameters and ability estimates from IRT.

As I mentioned earlier, most of the work on these indices has been statistical in nature. This work has established the fact that these indices are reasonably effective in detecting departures from expected response models. By contrast, very little work has been done on the determinants of such departures. Many possible sources of departure have been suggested, such as cheating, response coding errors, and fatigue. But little empirical work has been done to establish these factors as sources of departure from expected response models. In other words, the construct validity of these indices has not been firmly established. As a result, we know that these indices detect something, but we have no clear idea about what this something is.

The present paper attempts to address this issue of the construct validity of measures of inappropriateness. The first step is to treat inappropriateness as a construct by exploring its meaning and its implications. The second step is to discuss factors that might be expected to lead to inappropriate responses and build these factors into a model of inappropriateness. The third step is to begin testing the model to see if the determinants that are included in the model actually do have an impact on measures of inappropriateness. To this end, I begin with a discussion of the foundations of mental testing in general and test appropriateness in particular, and how early concerns over appropriateness led to modifications in the conceptual model used to describe item responses. I then explain how specific individual and item characteristics might combine to affect both the level of test scores and the appropriateness of test scores for certain people. Finally, I describe a test of parts of this model.

Foundations of test appropriateness

Although the history of mental testing in general can be traced back thousands of years to the ancient Chinese and Greeks (DuBois, 1966; Anastasi, 1988), the roots of contemporary testing can be found in the early nineteenth century. The work of Galton, Cattell, Binet, Terman, Goddard, and others is well documented and need not be reiterated here (see Hothersall, 1990 or Boring, 1950 for thorough reviews). One theme that runs through the work of all of these early testing experts is an assumption that any given mental test measures the same constructs (although the term "construct" wasn't used) for all people. In other words, item responses are determined only by individual differences on the trait of interest and, where appropriate, item difficulties. The possibility of test inappropriateness for certain individuals was not considered. One of the more striking examples of this assumption at work comes from the testing of immigrants at Ellis Island in 1914. At Ellis Island, immigrants were asked in their own language several trivia questions developed by Goddard and his staff. The trivia questions consisted, for the most part, of bits of Americana such as "Who is Christy Matthewson?" and "What is Crisco?" and were designed to assess intelligence. With the benefit of hindsight, it is obvious that these questions, while perhaps valid as measures of intelligence for an American sample, were utterly inappropriate for immigrants from Italy, Hungary, Russia, etc. In other words, these items assessed different constructs for different respondents depending on whether the respondents were American or not. This contamination, however, was not identified by Goddard. Because he assumed that the item responses were caused by the level of intelligence of the respondent and the difficulties of the items

and nothing more, he concluded that over 80% of immigrants to the United States were "feebleminded" (Hothersall, 1990).

Yerkes was among the first to identify individual differences other than the construct of interest that affect item responses. In the preliminary testing of the Army Alpha test of intelligence or "native wit" in 1917, he recognized that many of the respondents were not sufficiently literate to follow the instructions for the test (Hothersall, 1990). In other words, Yerkes recognized that individual differences other than native wit, namely reading skills, were determining item responses. For this reason, the Army Alpha test was inappropriate as a measure of intelligence for the illiterate. It was in response to this issue that the Army Beta was developed.

Cady (1923), Allport (1928), and Rosenzweig (1934) were among the first to identify item characteristics other than difficulty (or the personality-test equivalent of item difficulty, item popularity) that affect item responses. These authors suggested that item characteristics such as social desirability would also have an effect on item responses, at least for some respondents.

What I have presented above are the components of a very general model of mental test item responses. Item responses are determined by the respondent's level on the construct of interest, the degree to which various extraneous constructs affect the reaction of the respondent to items, the difficulty of the construct-relevant content of the item, and other construct-irrelevant factors that influence item responses. This model is presented in Figure 1.

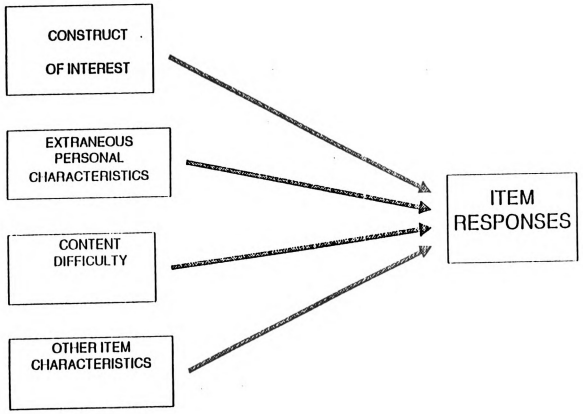

Figure 1. A general model of the determinants of item responses as they relate to test appropriateness

To the extent that extraneous respondent characteristics have an impact on item responses for a given respondent, the test composed of those items is inappropriate as a measure of the construct of interest for that person.

This conceptual model went largely unchanged until 1968 when Donlon & Fischer introduced their index of test appropriateness: the personal biserial coefficient. The specifics of the personal biserial are described later in this paper. The point that I wish to make about the personal biserial here is that it represents the first effort to identify threats to appropriateness in the form of interactions between item characteristics and characteristics of the respondent. Since the personal biserial is an index of the relationship between group-determined item difficulties (computed from a sufficiently large sample) and dichotomously scored item responses for a given respondent, it is an index of the extent to which the item difficulties hold for a given respondent. In other words, it assesses the strength of the relationship between item responses and the interaction between group-determined item difficulties and characteristics of the respondent. To the extent that the relationship between item characteristics and item responses depends on or covaries with the level of a characteristic of the respondent, the test composed of those items is measuring different constructs for different people, i.e., the test is inappropriate as a measure of the construct of interest.

If we include person by item interactions, the model in Figure 1 is modified into the model in Figure 2.

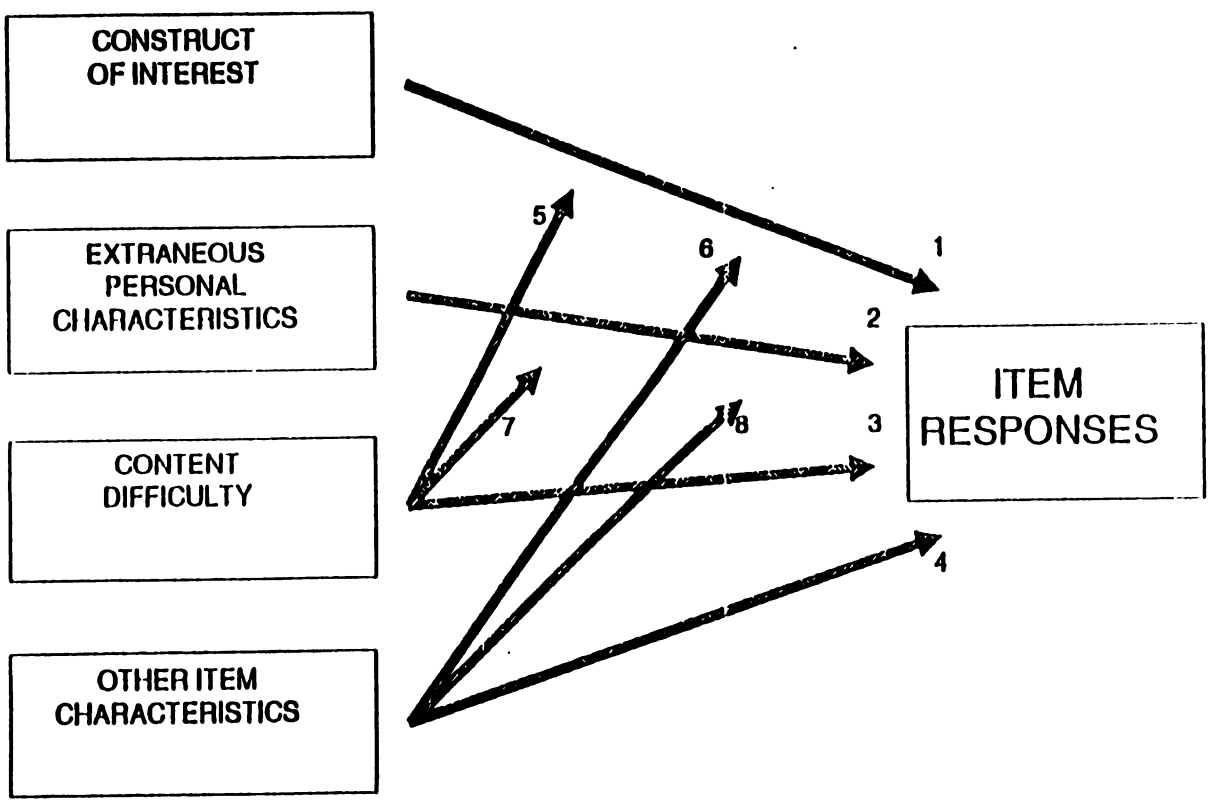

Figure 2. A general model of the determinants of item responses with interaction effects

Now, it would appear that we have two different types of sources of test inappropriateness. The first type, which I will call Type P (for Personal characteristics) inappropriateness, results from test items measuring characteristics of respondents other than those that the test was intended to measure. The second type, which I will call Type I (for Interaction) inappropriateness, results from an effect on item responses of the interaction between personal characteristics of respondents and item characteristics. What follows is a discussion of the various sources of these two types of inappropriateness followed by a discussion of the measures that have been used to identify them. The sources of inappropriateness that are discussed are the most prominent in the testing and measurement literature. Before moving on to this discussion, it should be noted that many of the sources that are discussed may seem more applicable to typical performance measures (e.g., personality tests) than to maximum measures. It is my position, however, that most of the sources of inappropriateness that may seem relevant only to one type of test have an analog in the other type. For example, response sets such as extreme response set (the tendency to give only extreme ratings), which are discussed in more detail below, may seem to be relevant only to personality-type tests, but they have an analog in maximum tests. The analog is what I refer to as positional bias or the tendency to choose the extreme response options when unsure of or uninterested in the correct response. So, many of these sources of inappropriateness can be conceptualized as relevant for either maximum or typical tests. An effort is made in the discussion that follows to point out both the maximum and typical sides of each of the sources of inappropriateness.

Sources of Type P inappropriateness

Acquiescence/denial. Acquiescence is the tendency of a respondent to uniformly agree with statements that are made. Denial is the tendency of a respondent to uniformly disagree. Although this applies primarily to typical tests where respondents are asked to agree or disagree, or provide a rating of the extent to which they agree, it could also apply to maximum tests with only true/false response options.

Acquiescence was originally investigated as a source of error in personality tests (Humm, Storment, & Iorns, 1939; Humm & Humm, 1944; Cronbach, 1946), and it is one of the few sources of inappropriateness that doesn't seem to have an analog in common multiple choice maximum tests.

Acquiescence, like all response sets, tends to emerge when a respondent finds an item to be ambiguous or difficult. In fact, Messick (1966) claimed that there are two types of acquiescence: that based on misunderstanding of items or carelessness and that based on personality.

Acquiescence, and not item content, has been found to account for much of the variance in responses to the MMPI (Jackson & Messick, 1958, 1965; Bock, Dicker, & Van Pelt, 1969), although some have claimed that acquiescence is not a problem (Block, 1965; Rorer, 1965).

Acquiescence leads to particular types of profiles in personality and interest tests that may or may not reflect the true nature of the respondent. To the extent that the MMPI and tests like it measure acquiescence in addition to or instead of the constructs that they are supposed to measure, those tests are inappropriate as measures of those constructs for those people who are high in acquiescence, although it should be noted that

acquiescence itself has been conceptualized as a personality construct of interest (Couch & Keniston, 1960; Wiggins, 1962; Messick, 1966).

Need for approval. Need for approval has been suggested as an individual difference characteristic that causes certain people to paint a favorable or socially desirable picture of themselves (Crowne & Marlowe, 1964). This characteristic would lead people to respond to items in a way that they feel depicts them not as they are, but as they would like to be and as they would like to be perceived by others. This need can be personality based or situationally based (e.g., taking a test for research purposes versus taking a test as part of a selection battery).

This leads us to the well-known phenomenon of "Faking good". It has long been known that people can and do fake good on a wide variety of tests (personality tests: Ruch, 1942; Green, 1951: interest batteries: Kingston, George, & Ewens, 1956; Gehman, 1957: Rorschach: Henry & Rotter, 1956: intelligence tests: Saupe, 1960). To the extent that Faking good is taking place for a given individual, the test is inappropriate as a measure of the construct of interest.

The maximum test analog to Faking Good is simply "Cheating". The principle is the same. Some maximum test takers wish to depict themselves as they would like to be instead of as they are. In response to this desire, they cheat. Obviously, to the extent that a maximum test reflects cheating, it is inappropriate as a measure of the construct of interest.

Extreme response set/central tendency. Some investigators have found evidence of Extreme response set (Berg & Collier, 1953; Cronbach, 1946, 1950) and its opposite, central tendency (Gaier, Lee, & McQuitty, 1953; Damarin, 1970). Extreme response set

exists when a respondent tends to choose the extreme response options (i.e., the first in a list of options or the last in a list of options) over other response options irrespective of item content or correctness of the extreme options. Central tendency exists when a respondent tends to choose the middle response options over the extreme response options. Although both extreme response set and central tendency in typical tests have clear analogs in maximum tests, the maximum and typical forms of this source of inappropriateness are treated separately here. In this paper, I refer to the typical form of these sources of Type P inappropriateness as response sets, and to the maximum form as response biases.

Response sets and biases, like acquiescence, tend to emerge when items are difficult or ambiguous. Extreme response set leads to particular types of profiles in personality and interest tests. Depending on the particular test being taken, extreme response bias can spuriously raise or lower scores on intelligence tests (Metfessel & Sax, 1957, 1958; Rapaport & Berg, 1955). Again, to the extent that this response set or bias affects the score of a given respondent, the score is a misrepresentation of the constructs of interest.

Test Anxiety. Although the concept of anxiety has existed in psychology since shortly after the inception of psychology, the notion of anxiety with respect to particular, normal situations is relatively new. Mandler & Sarason (1952) were among the first to discuss such a construct when they presented their measure of test anxiety. They found that test anxiety, by causing responses to the test situation that were irrelevant to cognitive test performance, decreased test performance. Mandler & Sarason (1952) and Waterhouse & Child (1953) also found an interaction between these situation-specific trait anxieties

instructions, etc.) such that those respondents who were low on the situation-specific trait anxiety seemed to benefit slightly from the general situational anxiety whereas those respondents who were high on the specific trait anxiety showed a decrease in performance.

Although there was some initial skepticism about the distinctiveness of specific trait anxieties such as test anxiety, there now seems to be reasonable agreement on its distinctiveness (Harper, 1976; Watson & Clark, 1984). Test anxious people have been found to have poorer study habits, fact retention, and elaborative processing, as well as diminished abilities to synthesize or analyze relevant information (Herrmann, 1982). They have been found to have deleterious cognitive responses to test situations (Endler & Hunt, 1966), and show deficiencies in all stages of information processing (Benjamin, McKeachie, Lin, & Holinger, 1981). Also, Naveh-Benjamin, McKeachie, & Lin (1987) distinguished between two types of test anxious people: those with good study habits who have trouble only with information retrieval in the test situation, and those with poor study habits who have trouble with all stages of information processing. The implication of Naveh-Benjamin et al. (1987) is that some respondents are test anxious because they are not prepared, while others have no good reason for being anxious. If this is the case, then the former type of anxiety would perhaps be better labelled something else, since it is not the test per se that these respondents are anxious about. Rather, they are anxious about a test that they are not prepared for, which would seem to be something entirely different from general anxiety in testing situations.

The implications of test anxiety for typical tests should be similar to those for maximum tests. Test anxiety inhibits information processing, which means that highly test anxious respondents should have more difficulty providing accurate responses to items on typical tests than do low test anxious respondents.

To the extent that a test reflects test anxiety or similar constructs (e.g., number anxiety; Dreger & Aiken, 1957; frustration anxiety; Waterhouse & Child, 1953), it is inappropriate as a measure of the construct of interest for that respondent.

Cognitive Controls. Cognitive controls are involuntary ways of approaching and interpreting complex situations or stimuli. They represent different ways in which we organize the information that we are constanly receiving from the external world. Among the various types of cognitive controls are field articulation, equivalence range, levelling-sharpening, cognitive complexity, and scanning. Although many researchers have hypothesized relationships between these individual controls and intellectual abilities (e.g. Gardner, Jackson, & Messick, 1960; Witkin, 1959; Klein, 1959; Gardner, 1959), there has never been a clear distinction among these various cognitive controls (McGee, 1979), and the construct validity of many of their measures has been called into question as well (Sherman, 1967; McGee, 1979). Of these cognitive controls, field articulation has received the most attention. It has been developed as a construct itself (Broverman, Klueter, Kobayashi, & Vogel, 1968), and measures of field articulation, such as the Witkin Embedded Figures Test, have reflected convergent and discriminant validity (Satterly, 1976; Witkin, 1974). For these reasons, this discussion focuses on field articulation to the exclusion of the other cognitive control principles.

Field articulation is the extent to which a person is able to pick out certain relevant aspects of a complex stimulus or situation to the exclusion of other, superfluous aspects. It has been shown to be related to scores on mathematics tests (Satterly, 1976), learning and memory (Goodenough, 1976), paragraph comprehension (Klein, 1967), and acquiescence (Forehand, 1962).

Field articulation might also be expected to affect responses to typical-test items. Specifically, respondents who are low in field articulation (also called field dependent respondents) may have difficulty extracting relevant information from questions on personality or interest inventories and may therefore have difficulty providing accurate information.

Response bias/test wiseness. Response bias refers to the tendency to select particular response options for reasons other than content. For example, Lawrence (1957) showed that children have a tendency to guess the first distractor presented in a list of distractors when unsure of the correct response. This tendency is a response bias. Extreme response bias, which was mentioned earlier, might be considered another example.

There is some confusion in the literature about the distinction between response bias and test wiseness (Sarnacki, 1979). Test wiseness has been defined as the capacity of the respondent to utilize the characteristics and formats of the test and/or test situation to receive a high score. The distinction between response bias and test wiseness seems to be a matter of rationale. If a respondent has a tendency to guess the extreme response options when in doubt of the correct answer, and the reason for this behavior is simply that the respondent is in the habit of guessing the extreme response options, then this

would be merely a response bias. If, on the other hand, the respondent had a tendency to guess the extreme response options because he/she had heard that this particular test-maker tended to put the correct options on the extreme positions, or if the respondent noticed that many of the correct responses on the early part of the test were in the extreme positions, and this were the reason for the reliance on the extreme options when guessing, then the use of the extreme positions would be an example of test wiseness. So, if the response tendencies of a respondent unsure of correct answers are based on whim or habit, they are response biases. If the response tendencies of a respondent unsure of correct answers are based on the characteristics of the test or testing situation, then they reflect test wiseness.

Test wiseness has been found to have an effect on a wide variety of multiple choice tests (Dolly & Williams, 1986; Wahlstrom & Boersma, 1968; Millman, Bishop, & Ebel, 1965). To the extent that a test measures test wiseness instead of the construct of interest, the test is inappropriate as a measure of the construct of interest.

Test wiseness can also affect scores on typical tests in certain situations. For example, if a mental health professional is given the MMPI as part of a battery of selection procedures, he/she would probably know enough about the test to be able to provide whichever type of profile that he/she wished. This knowledge of the MMPI would also carry over to other personality tests as well. In this way, the mental health professional would be using the characteristics of the test to his/her advantage.

The relationship between response bias and appropriateness is less clear. If response biases are exposed only when the respondent has no idea what the correct answer to an item is (or in the case of typical tests, what the most accurate answer is),

and if the response bias is in fact not based on rationale of any kind, then the response bias will affect item responses only in a random way, and there should be no effect on test scores. If, however, the response biases emerge even when the respondent does have some idea of the correct response, and if those biases override the content knowledge of the respondent such that the biases lead the respondent to choose an option that is at odds with his/her content knowledge, then the response biases would affect test scores.

Carelessness/motivation. One of the earliest identified contaminants of test scores was carelessness. Respondents who are not motivated to provide responses on a test that reflect that respondent's true standing on the construct of interest, be it maximum or typical, may respond carelessly to items. Also, respondents who are anxious about the test situation may accidentally provide answers that do not reflect his/her true standing on the construct of interest (e.g., coding errors on computer-scored answer sheets).

Peterson (1961) included nonsense items in an interest inventory as a carelessness check and found that many respondents endorsed them, especially when they were in the latter part of the inventory. This nonsense item technique has been incorporated into many noncognitive tests (e.g., MMPI, CPI). To the extent that a test reflects carelessness or motivation to perform to the best of one's abilities, the test is inappropriate as a measure of the construct of interest.

This brings us to a related topic: motivation. In real world testing situations, such as classroom testing and employee selection situations, tests are generally assumed to measure not one, but two constructs: the construct of interest (such as content knowledge), and motivation. Tests in real world contexts are developed primarily as measures of content knowledge and perhaps application, and though motivation can and

perhaps should be viewed as a contaminant, it is seldom examined or taken into account when test scores are interpreted. If motivation is a part of such tests, then it should be taken more seriously. Its effects on test scores should be examined more carefully, for a test that is designed to measure content knowledge or intelligence is inappropriate as a measure of those constructs to the extent that it is affected by motivation.

There is one final point that should be made with respect to motivation. I said earlier that motivation to provide responses on a test that reflect that respondent's true standing on the construct of interest was related to carelessness. One could make the argument that cheating or "faking good" would be another result of such motivation, but in the opposite direction. For the sake of simplicity, when I refer to motivation, I am speaking only of motivation (or lack of) that results in careless responding. Outcomes such as "faking good" have been/will be dealt with during discussions of need for approval.

Omissiveness. One final source of inappropriateness that should be mentioned is omissiveness. Omissiveness is the tendency to leave blank those items for which one is not sure of the correct or accurate answer instead of guessing. Although there is very little research on omissiveness, Rosenberg, Izard, & Hollander (1955) found that there was an "undecided" or omissiveness response set that affected responses to noncognitive (i.e., typical) test items. Also, Schuman & Kalton (1985) discussed research which showed that better educated respondents were less willing to endorse "Don't Know" response options on attitude surveys than were less educated respondents. To the extent that a test measures omissiveness instead of the construct of interest, the test is inappropriate as a measure of the construct of interest. It should also be noted that the

measurement of omissiveness is perfectly straightforward: it is assessed by counting the number of omissions. It is, therefore, not included in the section on measures of Type P inappropriateness.

Section summary. In the above section, several sources of Type P inappropriateness were identified. These sources were Acquiescence/Denial, Need for Approval, Extreme Response Set/Central Tendency, Test Anxiety, Cognitive Control, Response Bias/Test Wiseness, Carelessness, and Omissiveness. These factors have been found to affect scores on a variety of tests, both typical and maximum, and to the extent that those tests are intended to measure something other than the above-mentioned factors but are contaminated by one of these factors, those tests are inappropriate as measures of their respective constructs of interest. It should be noted that a certain amount of inappropriateness in a test is not a reason to do away with the test. The inappropriateness should simply be kept to a minimum.

Sources of Type I Inappropriateness

Type I inappropriateness is described above as being present to the extent that there is a person by item interaction that affects item responses. For example, Anderson (1990) suggests that the item responses of test anxious respondents (on a cognitive test) are more adversely affected by item difficulty than are respondents who are not test anxious, that is, although item difficulty affects the "correctness" of item responses for respondents who are low in test anxiety, it affects "correctness" more strongly for respondents who are high in test anxiety.

Although it would be possible to suggest any personal characteristic by item characteristic interaction as a source of inappropriateness, some are more plausible based

on previous research. It is these latter types of interactions that are the focus of this section. It should be noted that while most of the interactions discussed below have implications for both maximum and typical tests, some have implications for only one or the other. Where applicable, implications for both are discussed.

Before moving on to these Type I sources of inappropriateness, there is one final issue that must be addressed. I have thus far discussed personal characteristics that affect item responses, and I am about to discuss personal characteristic by item characteristic interactions that affect item responses. One might ask why I haven't discussed item characteristics separately as a source of inappropriateness. The reason is that item characteristics, if they have only a direct effect that is not moderated by personal characteristics, do not distinguish among people. In other words, item characteristics alone do not contribute to inappropriateness because, by acting alone, they affect every respondent in the same way. So, the most that item characteristics can do is to add a constant to the score of every respondent. Since this would not change our conclusions with respect to the standing of any of the respondents on the construct of interest, item characteristics alone cannot be considered sources of inappropriateness. It is only when their effects depend upon some personal characteristic of the respondent that they become relevant to appropriateness.

Test anxiety by item characteristics. One of the implications of the Anderson (1990) paper is that test anxiety should interact with any item characteristic that contributes to item difficulty. Although we often think of item difficulty solely in terms of the difficulty of the content of the item, item difficulty (i.e., the percentage of respondents answering the item correctly or, in Item Response Theory terms, the

probability of a respondent with a certain ability level answering the items correctly) is affected by a number of item characteristics.

Item difficulty has been found to be affected by ambiguity of item content (Peabody, 1966), item/stem complexity (e.g., word problems, Zimmerman, 1954), and response option complexity (e.g., none of the above, Hughes & Trimble, 1965; Dudycha & Carpenter, 1973). Research has also found that negatively worded items are more difficult than positively worded items (e.g., "Which of the following are not..." as opposed to "Which of the following are...", Dudycha & Carpenter, 1973), and that open stem items are more difficult than closed stem items (e.g., "Hitler was a member of the _____" as opposed to "Hitler was a member of which of the following parties?", Dudycha & Carpenter, 1973). Finally, there is evidence to suggest that item position can have an effect on item responses. For example, all else being equal, item to item transfer of training leads to later items being less difficult than earlier items (Whitcomb & Travers, 1957). In the typical performance test area, there is evidence that suggests that items later in a test are answered not on the basis of fact per se, but instead are answered in a way that will create consistency with responses to items appearing previously in the set of items (Schuman & Kalton, 1985; Feldman & Lynch, 1988).

In addition to these research findings, I suggest that one additional factor, topic irrelevant item content (e.g., reading component in a math test), contributes to item difficulty. It has been found that test anxiety is related to the complexity of a task through information processing (Benjamin et al., 1981; Naveh-Benjamin et al., 1987; Paulman & Kennally, 1984). Specifically, the cognitive component of test anxiety diverts cognitive resources from the task of test taking, thereby decreasing the amount of

cognitive resource devoted to the task. This suggests that any factor which increases the difficulty of a test item will interact with test anxiety to affect item responses. Specifically, I suggest the following proposition:

Proposition 1 - Test anxiety interacts with item content difficulty, ambiguity of item meaning, positive/negative stem wording, open/closed stem format, response option complexity, stem complexity, topic irrelevant item content, and item position to affect item responses such that the responses of test takers high in test anxiety are more adversely affected by these item characteristics than are the responses of test takers low in test anxiety. The form of this expected interaction is depicted in Figure 3.

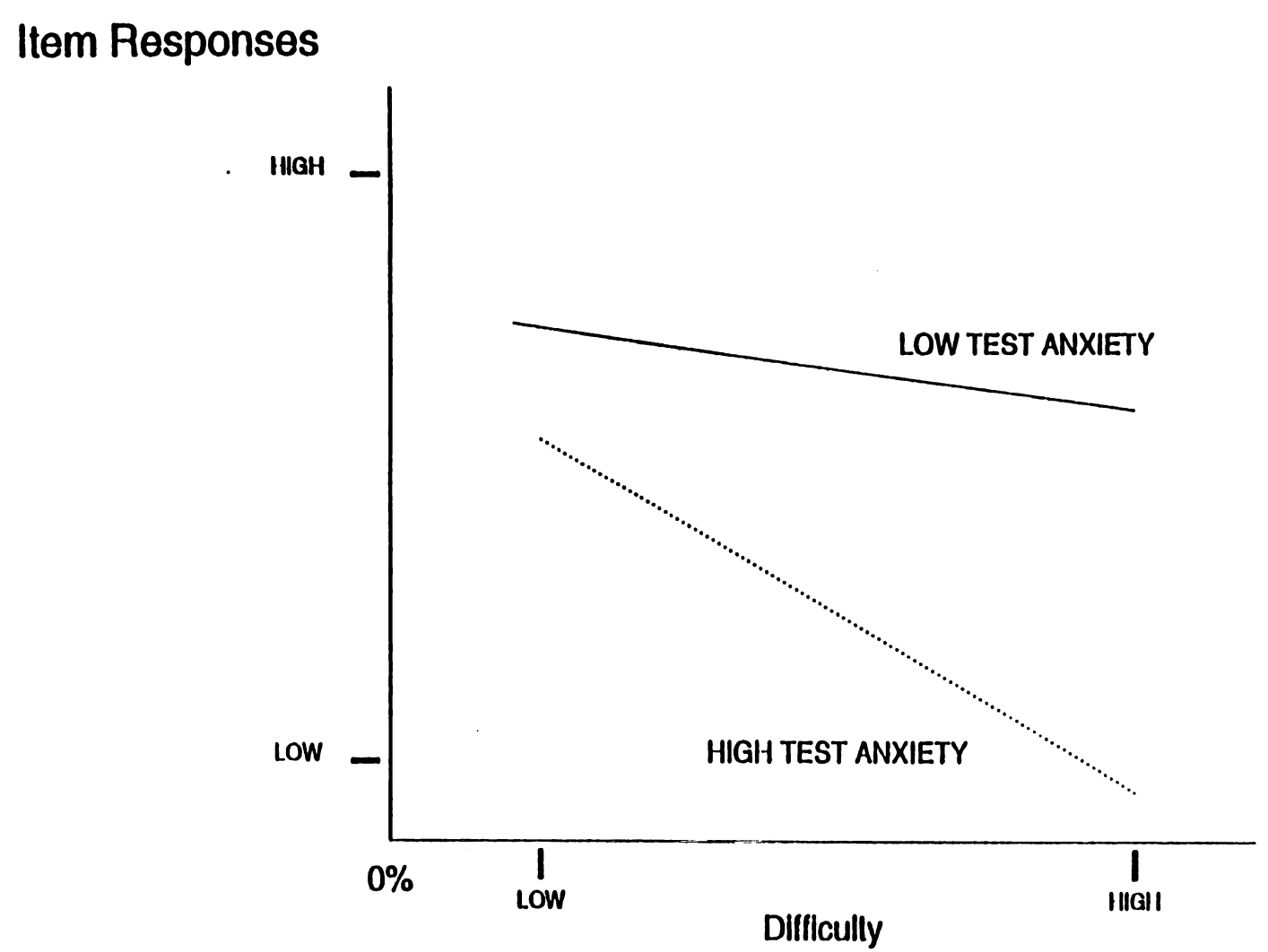

Figure 3. Proposed interaction between item characteristics and test anxiety

As can be seen, it is proposed that item difficulty in the form of various item construction characteristics has a slight, negative impact on item responses for low Test Anxious respondents and a considerably larger negative impact on the responses of high Test Anxious respondents. Although only one interaction is presented in Figure 3, it is intended to represent all of the item construction principles listed in Proposition 1.

All of these interactions, with the possible exception of that associated with content difficulty, could apply to both maximum and typical tests. For example, respondents high in test anxiety are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that are high in ambiguity than are respondents low in test anxiety, while this difference does not exist (or is less profound) for items low in ambiguity.

If this proposition is supported, then a test which contains items that vary with respect to any of the above mentioned characteristics is inappropriate (i.e., Type I) as a measure of the construct of interest to the extent that respondents vary with respect to test anxiety.

Motivation by item characteristics. It was mentioned earlier that a lack of motivation to provide responses that reflect one's true level on the construct of interest can lead to carelessness, which implies inappropriateness. Careless responding has often been identified as a contaminant of test scores in applied psychology. For example, one of the criticisms of concurrent validation designs is that incumbents, because they have nothing to gain by performing well on selection tests administered to them in a validation context, may respond carelessly to some items (Schmitt, Noe, Gooding, & Kirsch, 1984). It seems that they would be most likely to respond carelessly to those items that would

require more effort, i.e., the more difficult items. In other words, we would expect a motivation by item difficulty interaction. It has been shown that a variety of test takers are able to accurately estimate item difficulties, with correlations between true and estimated difficulty ranging from .56 to .77 (Diamond & Lorge, 1954). Likewise, any factor that contributes to item difficulty would be expected to interact with motivation to affect item responses. This suggests the following proposition:

Proposition 2 - Motivation interacts with item content difficulty, ambiguity of item meaning, positive/negative stem wording, open/closed stem format, response option complexity, stem complexity, and topic irrelevant item content to affect item responses such that respondents who are low in motivation reflect more carelessness on items that possess characteristics such as item content difficulty, negative wording, and complex response options than they do on items that do not possess these characteristics (e.g., items with simple content, positive wording, and simple response options). Respondents high in motivation reflect little carelessness in either case. The nature of this interaction is depicted in Figure 4.

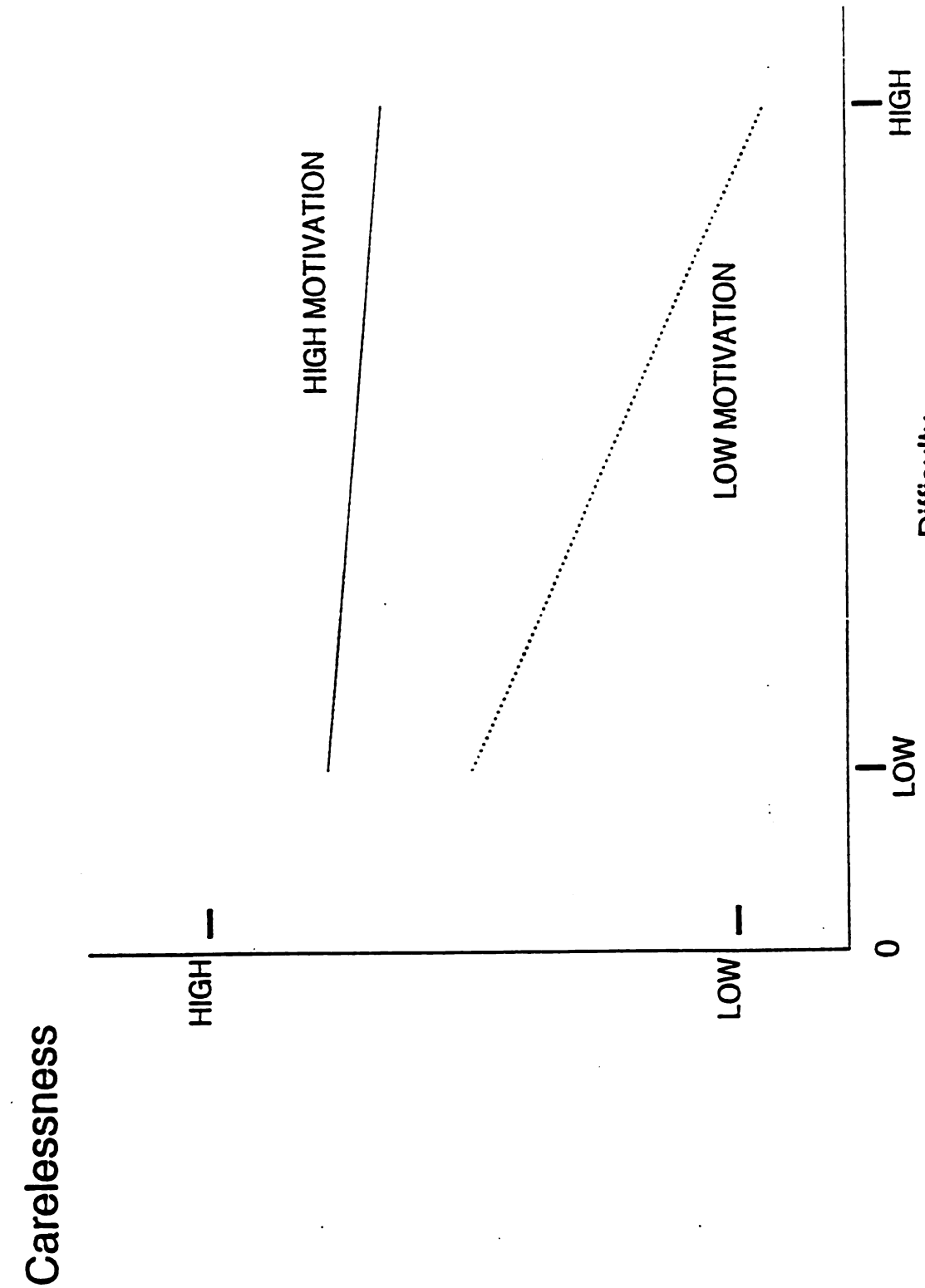

Figure 4. Proposed interaction between item characteristics and motivation

The form of the interaction presented in Figure 4 is similar to that of the interaction presented in Figure 3. For highly motivated test takers, item construction-based difficulty is expected to have a slight negative impact on item responses. For low motivation test takers, this negative effect should be more pronounced. Again, the form of this interaction holds for all of the principles listed in Proposition 2.

All of these interactions, with the possible exception of content difficulty, could apply to both maximum and typical tests. For example, respondents low in motivation are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that are negatively worded than are respondents high in motivation, while this difference does not exist (or is less profound) for items that are positively worded.

Motivation is not expected to interact with item position because, while it might be expected that the difficulty associated with items early in a test would affect low motivation respondents more severely than it would respondents high in motivation, fatigue effects over time would be expected to show similar effects for the two motivation groups, so that the position effects would be washed out.

Omissiveness by item characteristics. It was mentioned earlier that respondents vary with respect to their reactions to items when they are unsure of the answer. Some respondents will guess at any item even if they have no idea of the correct response (what Cronbach (1946) called the gambler's mentality), while others will tend to leave such items blank. Omissiveness may not be common among the generation of school-goers that grew up with standardized tests, since these respondents would generally know to guess at every item unless told to do otherwise. The older generation, however, along

with the less-educated, are less likely to possess such test wiseness. These respondents may reason that they should leave items for which they have no response blank since they should, in fact, get the item wrong. Since it is quite possible to guess correctly on multiple choice tests, differences in omissiveness will lead to differences in test/item scores. Furthermore, if respondents are more likely to guess or fail to guess at the more difficult items, any factor that contributes to item difficulty should contribute to the omissiveness by item interaction. This suggests the following proposition:

Proposition 3 - Omissiveness interacts with ambiguity of item meaning, difficulty of item content, positive/negative stem wording, open/closed stem format, response option complexity, stem complexity, topic irrelevant item content, and item position to affect item responses such that the responses of respondents high in omissiveness are more adversely affected by these item characteristics than are the responses of respondents low in omissiveness. In typical test terms, respondents high in omissiveness are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that possess characteristics such as stem complexity and topic irrelevant content than are respondents low in omissiveness, while this difference does not exist (or is less profound) for items that do not possess these characteristics. The nature of these interactions should be similar to that presented in Figure 3 and will therefore not be depicted graphically.

Field articulation by item chacteristics. It was said earlier that cognitive controls are involuntary ways of approaching and interpreting complex situations or stimuli, and that one of these controls, field articulation, is the extent to which a person is able to pick out certain relevant aspects of a complex stimulus or situation to the exclusion of other, superfluous aspects. If respondents vary with respect to field articulation, then field articulation by item characteristic interactions would be possible for those item characteristics that serve to distinguish among respondents with different levels of field articulation. In particular, those item characteristics that contribute to the complexity of the item/stimulus should interact with field articulation to affect item responses. For example, it might be expected that respondents low in field articulation (i.e., field dependent respondents) would have more difficulty with word problems (i.e., items embedded in a context) than would respondents high in field articulation. This suggests the following proposition:

> Proposition 4 - Field articulation interacts with item stem complexity, topic irrelevant item content, and response option complexity to affect item responses such that the responses of respondents low in field articulation are more adversely affected by these item characteristics than are the responses of those persons high in field articulation. In typical test terms, respondents low in field articulation are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that possess these characteristics than are respondents high in field articulation, while this difference does not exist (or is less profound) for items that do not possess these characteristics (e.g., items with

simple stems and no topic irrelevant content). The nature of these interactions is also expected to be similar to that presented in Figure 3 and will therefore not be depicted graphically here.

Response bias by susceptibility to bias. It was said earlier that response biases are response tendencies (such as central tendency, extreme response bias, etc.) of a respondent unsure of correct answers based on whim or habit, and that if correct response options were evenly distributed about the response positions, then the response bias will affect item responses only in a random way, and there should be no effect on test scores. Research has shown, however, that correct responses often are not evenly distributed (Metfessel & Sax, 1957, 1958). Therefore, a given response bias, although whimsical, can lead to higher or lower test scores if the items in the test are susceptible to such bias. This suggests the following proposition:

Proposition 5 - Response bias interacts with susceptibility of a test to response bias to affect item responses such that respondents who possess a particular response bias receive test scores that are higher than those of respondents who do not possess the bias if the test is loaded with items whose correct options are in positions that are likely to be chosen by the respondent with the bias. If the test is loaded with items whose correct options are in positions that are not likely to be chosen by the respondent with a particular bias (e.g., many items with correct answers in the extreme options given to a respondent with a central tendency bias), then that respondent receives a lower test score than the respondent without the bias. In typical test terms, a respondent who possesses a particular response

set is less likely to provide responses that accurately reflect her/his true standing on the construct of interest on a test that is loaded with items whose options that are "correct" for that person (i.e., that best reflect the respondent's true standing on the construct of interest) are in positions that are unlikely to be chosen by the respondent with the particular bias than is a respondent who does not possess the set, while this difference does not exist (or is less profound) on a test that is not loaded with such items. The nature of this interaction is presented in Figure 5.

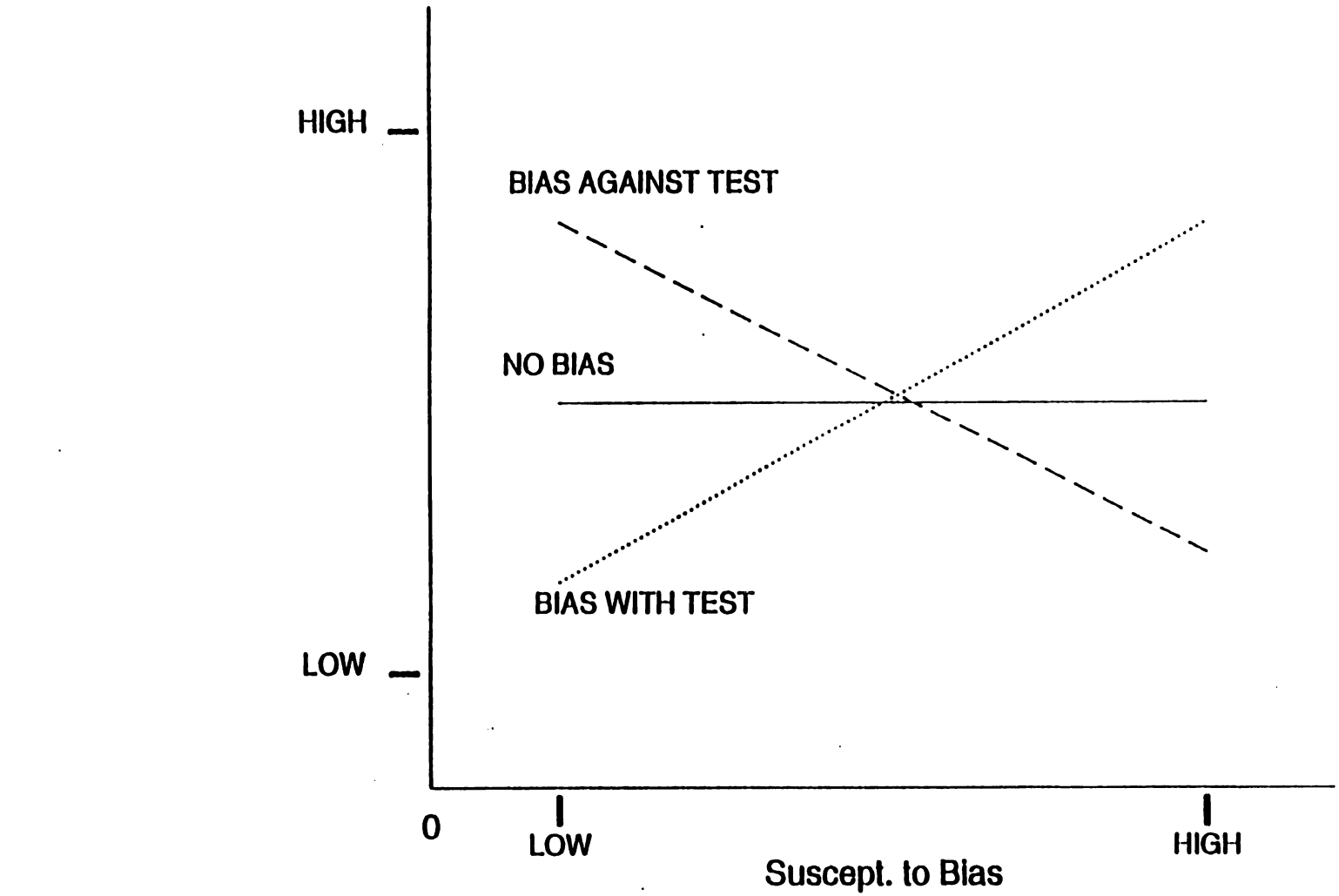

Figure 5. Proposed interaction between response bias and susceptibility of items to bias

Figure 5 suggests first that, for respondents with no response biases, susceptibility of test items to bias has no effect on item responses. For those test takers whose response bias is contradictory to the susceptibility of the items (i.e., a respondent with a predilection for extreme option positions who takes a test with a preponderance of correct options in the middle positions), susceptibility should have a negative effect on item responses, while the opposite should occur for respondents whose bias is in line with the bias of the test.

Test wiseness by susceptibility to wiseness. It was said earlier that if the response tendencies of a respondent unsure of correct answers are based on the characteristics of the test or testing situation, then they reflect test wiseness. Any characteristic of items that tends to elicit this test wiseness in those respondents who possess some degree of test wiseness should interact with wiseness to affect item responses. For example, consider the following item from a test of metric system knowledge (assume that respondents are instructed to select the single "best" answer to each question):

Which of the following is the unit of measurement closest to a unit in the English system?

a. one-thousandth of a liter

b. one milliliter

c. one centiliter

d. one decaliter

Now consider the possible answers of two respondents of equal ability, one of whom is high in test wiseness, the other low, with neither possessing the content knowledge necessary to answer this question. The respondent low in test wiseness has no recourse but to guess blindly, with a probability of a correct response equal to .25. The respondent high in test wiseness, however, can use the fact that options a and b are equivalent to discard both of them as possible correct responses. Thus, this respondent has only two options to guess from, with a probability of correct response equal to .50. It can be said that this item is susceptible to test wiseness because one of the more commonly identified aspects of test wiseness, deduction, can be used to increase one's chances of responding correctly to the item. If response b had been "one liter" instead of "one milliliter", then the susceptibility to test wiseness of the item would be removed, and the probabilities of correct responses for the two respondents would be equal. This suggests the following proposition:

> Proposition 6 - Test wiseness interacts with the susceptibility of items to test wiseness to affect item responses such that test wiseness leads to higher probabilities of correct responses on items that are susceptible to test wiseness, but is unrelated to the probability of correct response for those items that are not susceptible to test wiseness. To the extent that some form of benefit accrues to respondents with particular profiles on typical tests (e.g., a personality test used as a selection instrument), this interaction is expected to hold for typical tests as well. This interaction is presented in Figure 6.

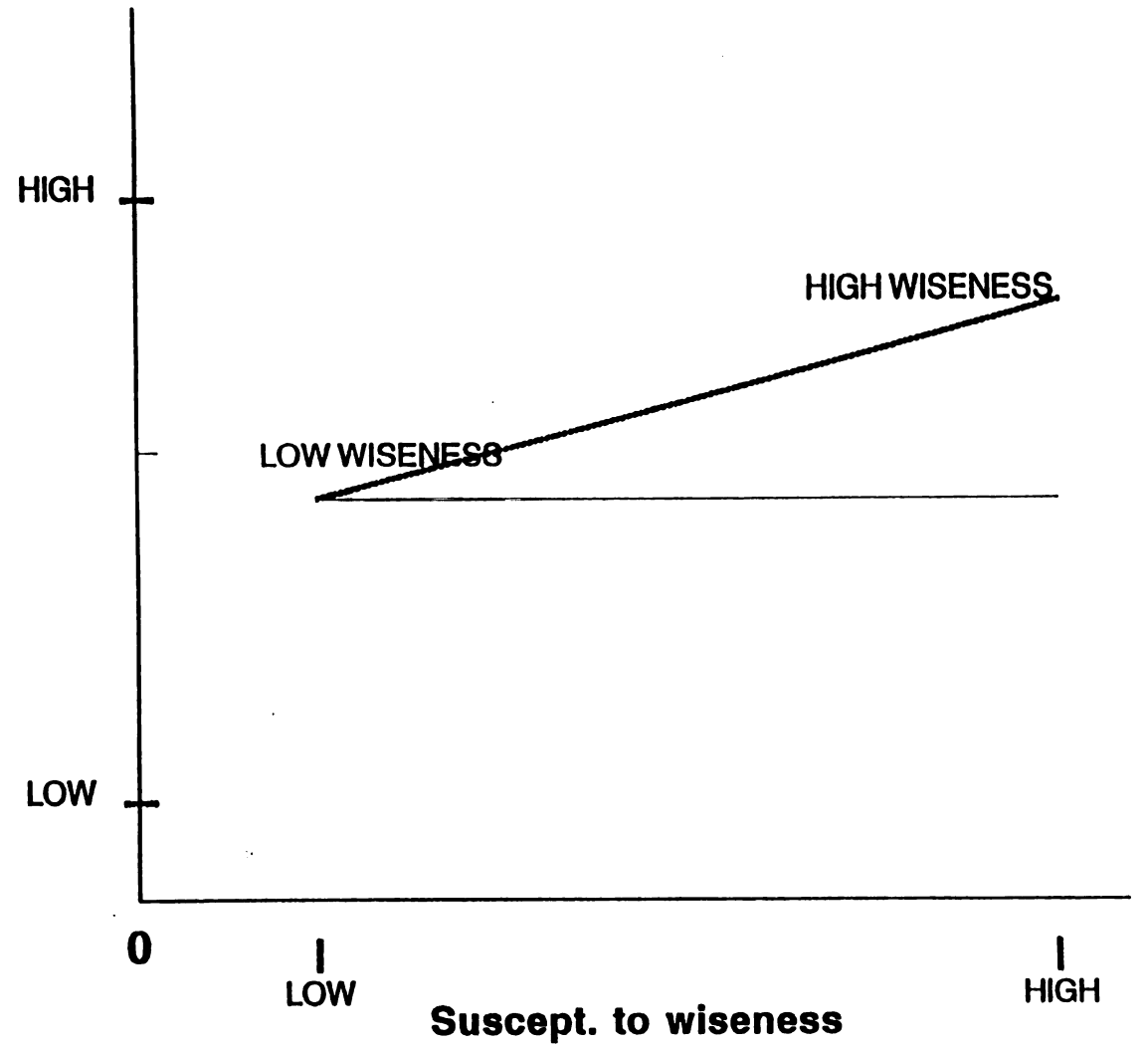

Figure 6. Proposed interaction between test wiseness and susceptibility of items to test wiseness

As can be seen in Figure 6, the proposed relationships among wiseness, susceptibility of items to wiseness, and item responses are identical to those among bias with the test, susceptibility to bias, and item responses. Specifically, test wiseness is beneficial to the "test wise" on items that are susceptible to wiseness and has no effect on items that are not.

Topic irrelevant ability by topic irrelevant item content. It was discussed above that a test is inappropriate as a measure of the construct of interest to the extent that it measures a construct other than the construct(s) of interest. Any item characteristic which affects the relationship between the topic irrelevant ability of a respondent (e.g., verbal ability on a math test) and item responses can be said to moderate that relationship. For example, there is no reason to suspect that verbal ability would have an effect on a respondent's answer to a calculation problem in mathematics (i.e., a problem with no words, just numbers, such as 2 X 4). There is, however, reason to suspect that verbal ability would have an effect on a respondent's answer to a verbally phrased math problem (e.g, What is the product of two and four?). This suggests the following proposition:

Proposition 7 - Topic irrelevant ability interacts with topic irrelevant item content to affect item responses such that the responses of respondents low in topic irrelevant ability are more adversely affected by topic irrelevant item content than are the responses of respondents high in topic irrelevant ability. In typical test terms, respondents low in standing on the irrelevant construct are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that are high in topic irrelevant content (i.e., content that matches

the irrelevant construct than are respondents high in the irrelevant construct, while this difference does not exist (or is less profound) for items low in topic irrelevant content. This interaction is presented in Figure 7.

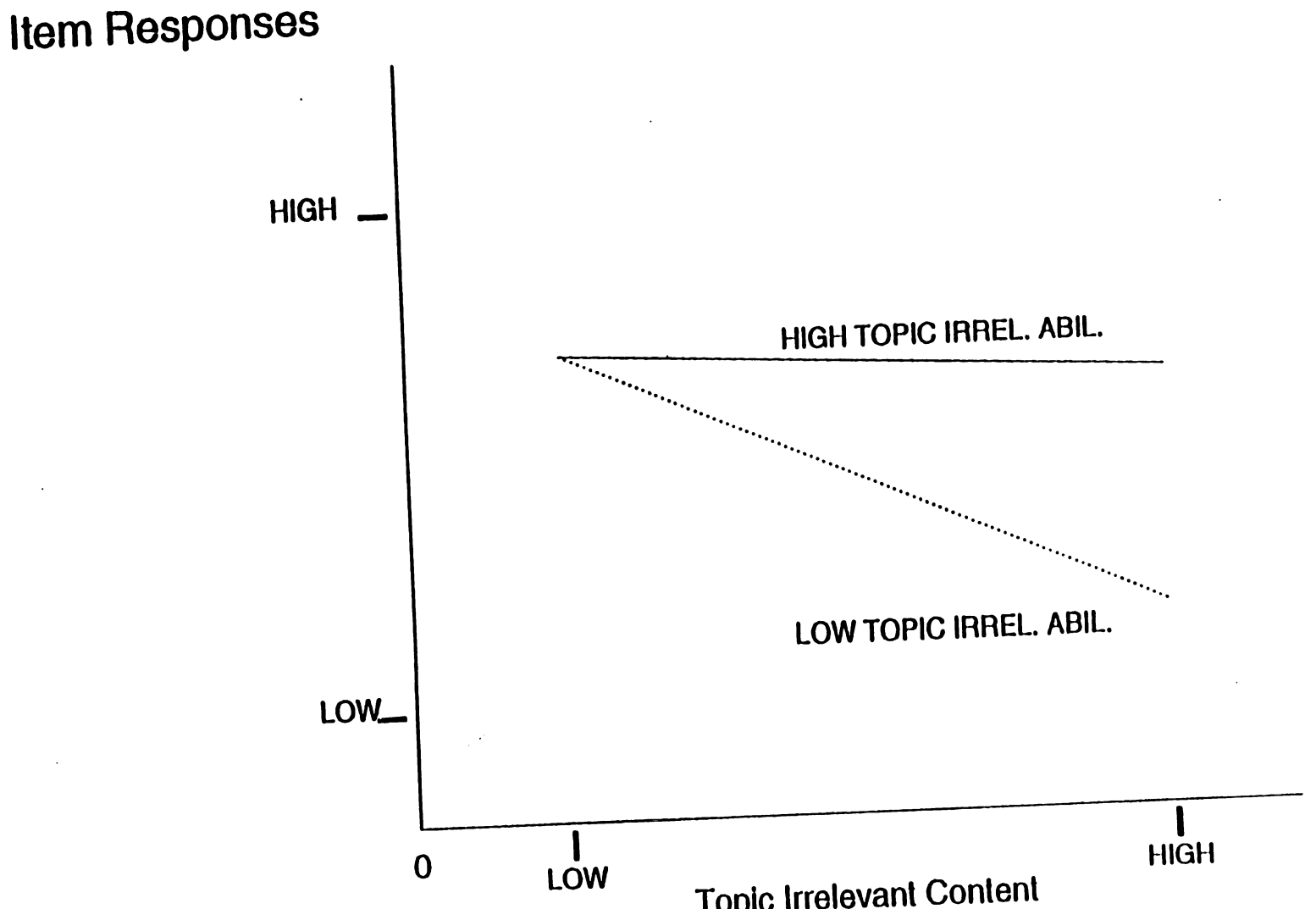

Figure 7. Proposed interaction between topic irrelevant ability and topic irrelevant content

Figure 7 shows that topic irrelevant content may have a slight negative impact on item responses for test takers high in topic irrelevant ability, but a large negative impact on responses of test takers low in topic irrelevant ability.

Need for approval by item characteristics. The final form of Type I inappropriateness to be discussed here involves need for approval. Specifically, certain types of items are more likely than others to elicit responses that reflect the need for approval of the respondent (This body of literature is discussed in more detail in the section titled, "Measures of Type P Inappropriateness). While need for approval is usually studied as one potential determinant of responses to typical test items, the responses to maximum test items that are the result of cheating seem to be psychologically equivalent to responses on typical test items that are the result of need for approval (as it has been conceptualized). In the maximum domain, certain items on a test might be more susceptible to cheating than others. For example, items at the ends of columns on bubble sheets might be easier to identify for someone copying answers than would items surrounded on all sides by other items. In the typical domain, items have been found to vary with respect to social desirability (Marlowe & Crowne, 1964). For both maximum and typical tests, items may vary in the extent to which they reflect need for approval. This suggests the following proposition:

Proposition 8 - Need for approval interacts with the characteristics of an item that might serve to reflect such a disposition to affect item responses. In the maximum domain, items which provide more of an opportunity to cheat result in inflated scores for respondents who are high in need for approval but not for

respondents low in need for approval. Items which provide little or no opportunity to cheat reflect no such difference across respondents. In the typical domain, respondents high in need for approval are less likely to provide responses that accurately reflect their true standing on the construct of interest on items that are high in social desirability than are respondents low in need for approval, while this difference does not exist (or is less profound) for items low in social desirability. This interaction is presented in Figure 8.

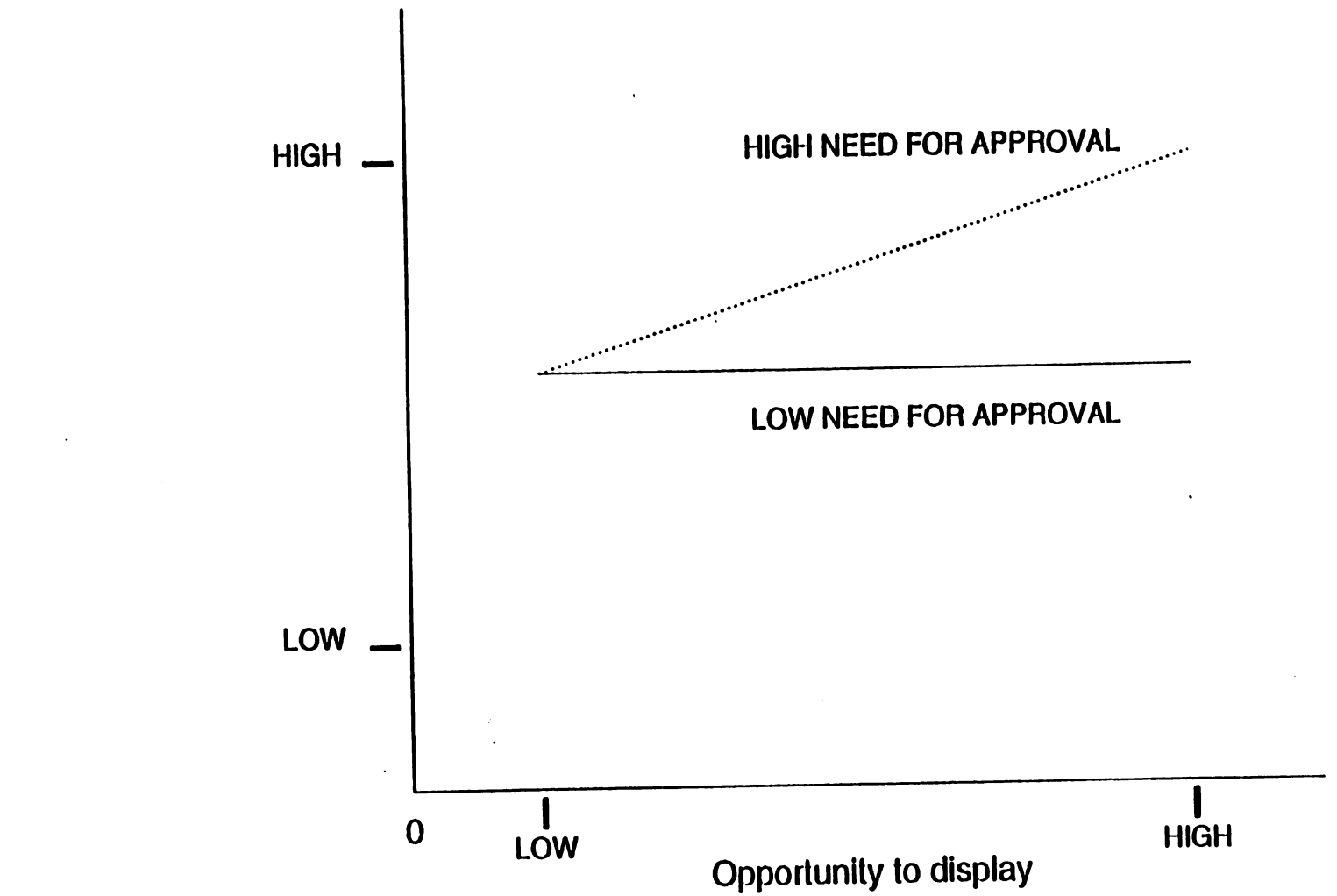

Figure 8. Proposed interaction between need for approval and opportunity to display need for approval

Figure 8 shows that opportunity to display need for approval, whether the opportunity be a clear view of a neighbor's paper or an item high in social desirability on a typical test, has no effect on the responses of test takers low in need for approval but a large positive impact on the responses of test takers high in need for approval.

Section summary. In this section, I reviewed some of the forms that Type I inappropriateness can take and suggested specific propositions about the form of interactions between personal characteristics of respondents and item characteristics. If these propositions are mapped onto the model presented on page 6 along with the sources of Type P inappropriateness discussed earlier in this paper, a new model (Figure 9) emerges.

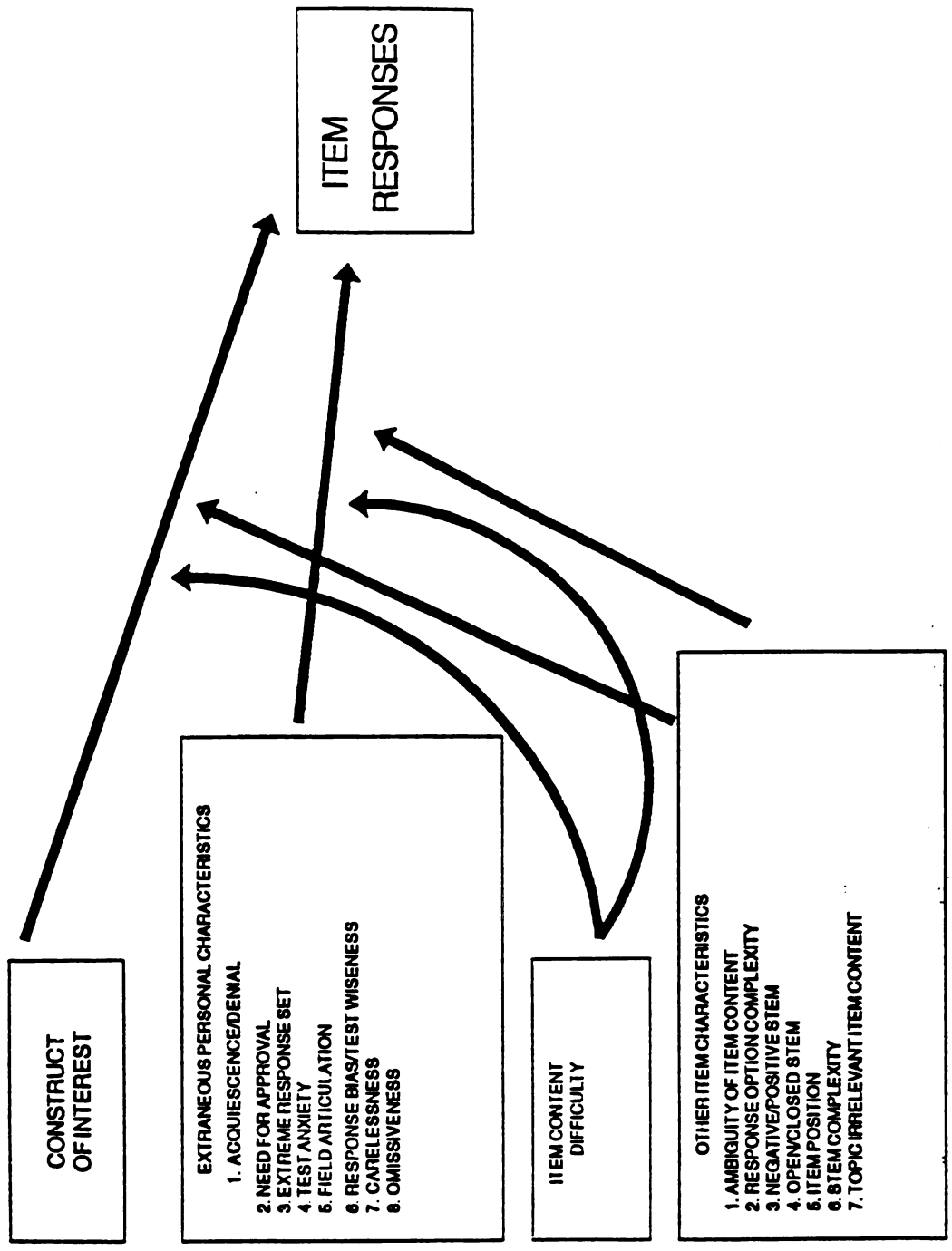

Figure 9. A detailed model of the determinants of item responses as they relate to appropriateness

In this model, all personal characteristics of respondents have direct effects on item responses, and all of these except the impact of the standing of the respondent on the construct of interest represent Type P inappropriateness. Also, many of the item characteristics moderate the effects of the personal characteristics, and these moderating effects represent Type I inappropriateness. The next section of this paper deals with the measurement of Type P and Type I inappropriateness.

Measures of Type P inappropriateness

The sources of Type P inappropriateness have direct, simple effects on test scores. They can be studied as main effects. As such, they can often be measured directly. I now describe the various ways that sources of Type P inappropriateness have been measured. Where relevant, I discuss differences between measures for Maximum and typical tests.

Acquiescence/denial and extreme response set/bias. Since these two sources of inappropriateness have been measured in similar ways, they will be treated together. The general method for assessing acquiescence/denial or extreme response set and their effects on test scores (both Maximum and typical) has been to compare the number of response options in question that were endorsed (e.g., the number of "agree" responses for acquiescence, the number of extreme options for extreme response set) to that expected by chance (Humm & Humm, 1944; Jackson & Messick, 1958). The problem with this method is that deviation from a chance model may simply reflect the actual standing of the respondent on the construct of interest. The solution to this problem for acquiescence was to reverse the wording of some items such that a person could contradict him/herself by agreeing categorically to all items.

Although no such solution has been devised for extreme response set, it is generally accepted that extreme response set, because it reflects an exaggeration of the true level of the respondent on the construct of interest instead of a complete distortion (as is the case with acquiescence), is not as serious a problem (Cronbach, 1950).

Need for approval. One of the earliest response sets to be identified was the tendency to "Fake good" on personality tests (Ruch, 1942), interest batteries (Gehman, 1957), and even projective tests (Henry & Rotter, 1956). This tendency was often studied but little understood until the work of Crowne and Marlowe (1964). These authors linked the tendency to Fake good to an involuntary need for approval that led certain respondents to display themselves in as favorable a light as possible. Their measure of need for approval (the Crowne-Marlowe Social Desirability Scale), which has been incorporated into many of the more prominent personality tests, such as the MMPI, involves True-False questions that reflect large amounts of social desirability but which should be answered in only one direction by virtually all respondents who are responding honestly. As Crowne & Marlowe aptly describe such items, " First, they are "good", culturally sanctioned things to say about oneself, and second, they are probably untrue of most people (p.210.)." Also included are items which are undesirable but probably true. An example of the former type would be, "Before voting, I thoroughly investigate the qualifications of all the candidates." An example of the latter would be, "I sometimes feel resentful when I don't get my way." Although the first item contains a most admirable quality in a person, it is assumed to be false for the vast majority of respondents who are responding honestly. Likewise, although the second item may not be admirable, it is probably true of most people. In this way, the Crowne-Marlowe Scale

and others like it (e.g., MMPI F-Scale; Edwards, 1957; Hartshorne & May, 1928) seek to identify those people who are attempting to provide a profile of themselves that is "socially desirable" instead of accurate.

The question that remains is, What do we do once we have identified a person who may be responding in this fashion? One approach has been to retest them in an attempt to get better measures of the constructs of interest. This approach can be used for both typical and Maximum tests. A second approach for typical tests has been to try to correct scale scores based on Social Desirability scores.

Test Anxiety. Although test anxiety has existed as a concept in psychology for at least forty years, it has been measured almost exclusively with the Mandler & Sarason (1952) measure, which has withstood much scrutiny (see discussion above on Test Anxiety). Nevertheless, Morris, Davis, & Hutchings (1981) developed a measure of test anxiety which appears to improve upon the Mandler & Sarason (1952) measure by tapping both the cognitive and emotional aspects of test anxiety. When test anxiety is identified as having an impact on a given individual's test score, several methods for decreasing the anxiety can be employed. For example, test anxiety has been shown to decrease as a function of instructional method (Tobias, 1979), study habits training (Naveh-Benjamin et al., 1987), feedback (Campeau, 1968), and training in positive affective responses (Watson & Clark, 1984).

Cognitive controls. Because there are several different cognitive controls that have been identified in the literature, there are dozens of cognitive control measures. The present study focusses solely on field articulation. For this reason, only measures of field articulation will be considered. The two most commonly used measures of field

articulation are the Embedded Figures Test and the Rod and Frame test. Both tests involve the identification of a figure of some kind that is embedded in a larger visual context. Thus, the high field articulation individual can sort through the irrelevant, contextual features and identify the figure. The low field articulation individual (or field dependent individual) has difficulty separating figure from context.

One additional measure of field articulation was discussed in Broverman et al. (1968). These authors explained sex differences in field articulation with certain neurological differences that are, in turn, caused by hormonal differences between the sexes. Although these neurological differences could, perhaps, be used as measures of field articulation, there has been no such attempt reported in the literature.

Although there is a small amount of research which suggests that field articulation does respond to training (Klein, 1967), it is difficult to assess the extent to which such training would really be helpful in sorting these effects out of scores on tests designed to measure other constructs. One option would be to partial out scores on tests such as the Embedded Figures from scores on tests of interest. Another option would be to eliminate items that are likely to contain a field articulation component, such as word problems. This second option, however, involves the interaction between field articulation and item characteristics, and will therefore be dealt with in the section on Type I inappropriateness.

Response Bias/Test wiseness. The distinction made earlier between response bias and test wiseness involves the rationale behind one's response strategy. Although the results of such a rationale (or lack of) can sometimes be identified through analyses similar to those used to identify extreme response set (cf., Fagley, 1987; Lawrence, 1957; Gaier et al., 1953), the rationale has been identified only with measures that are external

to the test of interest. One such measure has been that of Gibb (1964). This measure simply asks a respondent about the test strategies that he/she uses when taking a test, such as time-using strategies, error-avoidance strategies, guessing strategies, deductive reasoning strategies, estimation of instructor intent, and cue usage. Sarnacki (1979) used this measure to identify individual differences with respect to many of these strategies.

Carelessness. There are a variety of carelessness measures, but most of them have a similar form. Such measures contain items which, in one way or another, can be considered nonsense for the respondent. The nonsense content suggests that every respondent who is paying attention should respond in a particular way. For example, items from the MMPI K-scale or the Comrey Validity Check Scale might ask the question, "Have you ever been to the movies? Yes, Not sure, No" (Nonrandom Response Scale, Hough et al., 1990), to which, it is assumed, anyone who is paying attention should respond Yes. Such measures should "catch" any respondent who is responding carelessly regardless of the reason for the carelessness, be it lack of motivation, miscoding of responses, misunderstanding of the question, etc.

Carelessness, insofar as it is a function of the motivation of the test taker, can also be manipulated indirectly by manipulating motivation. The higher the motivation of the respondent, the less carelessness the respondent will exhibit.

Section summary. In this section, measures of the sources of Type P inappropriateness were briefly reviewed. These measures fall into two categories. The first category consists of those measures that are separate from the tests or items of interest. Examples are the Crowne-Marlowe Social Desirability Scale and the "Lie" scales

of the MMPI. For the most part, the measures of Need for approval, test anxiety, field articulation, test wiseness, and carelessness fall into this category.

The second category consists of those measures that result from reanalysis of data from the tests or items of interest. For example, extreme response set is typically measured by comparing the position of one's item responses to a chance model. There is no separate measure involved. For the most part, the measures of Acquiescence/denial, extreme response set/central tendency, and response bias fall into this category.

Measures of Type I inappropriateness

Because Type I sources of inappropriateness are interactions, they are more subtle than Type P sources and, therefore, more difficult to detect than Type P sources. As a result, the measures of Type I inappropriateness must also be more subtle, or at least more complex. The measures must be able to detect changes in the effects of personal characteristics on item responses that are due to changes in item characteristics. No measures with these properties have been recognized, but such measures do exist, they simply haven't been recognized.

These measures can be divided into two groups: those based on Item Response Theory and those based directly upon the pattern of right and wrong answers (Harnisch & Linn, 1981). These groups can be further divided into those for which some attempt to standardize has been made and those for which no such attempt has been made. More accurately, it can be said that both IRT-based and non-IRT based indices vary with respect to the extent to which they have been standardized relative to the total score (or theta) of the respondent. The meanings of these groupings are discussed in the sections

that follow. For more thorough reviews, see Harnisch & Linn (1981), Rudner (1983), and Birenbaum (1985).

Non-IRT based indices of Type I inappropriateness (unstandardized). One example of an unstandardized, non-IRT based index of Type I inappropriateness is Sato's Caution Index (1975). The formula for Sato's index is

$$C_i = \frac{\sum_{j=1}^{n_i}(1-u_{ij})n_j - \sum_{j=n_i+1}^{J} u_{ij}n_j}{\sum_{j=1}^{n_i} n_j - n_i\left(\frac{\sum_{j=1}^{J} n_j}{J}\right)}$$

where

i = 1,2,...I, indexes the examinee

j = 1,2,...J, indexes the item

u_{ij} = 1 if examinee i answers item j correctly and 0 if examinee i answers item j incorrectly

$n_{i.}$ = total correct for the ith examinee

$n_{.j}$ = total number of correct responses to the jth item

The name of the index comes from the idea that a large value indicates an unusual response pattern and, therefore, that caution should be used in interpreting the total score of this respondent (Harnisch & Linn, 1981). There are other such non-IRT based indices of Type I inappropriateness (e.g., the agreement/disagreement indices and the dependability index of Kane & Brennan, 1980; van der Flier's U (van der Flier, 1977) and its equivalent, the Nonconformity index of Tatsuoka & Tatsuoka (1980)), but they are

generally highly correlated with one another (Harnisch & Linn, 1981; Rudner, 1983) and the rationale is similar for all of them. Essentially, these indices answer the question, To what extent do the item responses of a given respondent conform to the item difficulties (as calculated from the total sample of examinees)? In terms of Type I inappropriateness, these indices answer the question, To what extent is there something about this respondent that renders the item difficulties invalid for this respondent? In other words, To what extent is there an interaction between the personal characteristics of the respondent and characteristics of the items? This question applies to both Maximum and typical tests. The only difference is that the notion of item difficulty in Maximum tests should be replaced in typical tests with some measure of the percentage of the sample endorsing a given response option.

Another way of describing the logic of these indices (as well as the IRT-based indices) is in terms of Guttman vectors. A Guttman vector is simply a vector of zeroes and ones in which all of the ones precede all of the zeroes. If a respondent were to respond to items in a way that matched perfectly with the difficulties of the items (and if there were no possibility of guessing), then the responses of that respondent, when ordered in terms of item difficulty, would form a perfect Guttman vector. The idea is that the respondent answers all items at or below a certain difficulty level correctly. At some point, however, the difficulty becomes too great for that respondent, and all items above that level of difficulty are answered incorrectly. If this were the case, then this respondent should receive a perfect score on the appropriateness index.

One reason for a departure from this perfect Guttman vector is that a respondent is guessing some items correctly. In a four option, multiple choice test, we would expect

a respondent to guess correctly 25% of the items that are too difficult for them. A second reason for a departure from a perfect Guttman vector is an interaction between personal and item characteristics. Consider the following. Item difficulties are calculated on an entire sample of scores. If the responses of a given respondent do not conform to those difficulties (for reasons other than guessing), then there is some characteristic of that individual respondent that is giving that person an advantage over the group on some items and/or a disadvantage over the group on other items, with the result being that the person answers correctly some items that should be too difficult for that person while answering incorrectly some of the easier items. The result is a vector of item responses (ordered by difficulty) like the following:

11111111100001111010

On the one hand, it would appear that the items became too difficult for this respondent after the ninth item in this order. On the other hand, this respondent did very well on the last seven items, items that the sample on which the difficulties were based found to be most difficult. There are two possible explanations (other than guessing). The first is that the content of the items beyond the ninth item was too difficult for this respondent, but something about this respondent (e.g., possession of a cheat sheet, a quick view to a neighbor's paper) gave him/her an advantage over the rest of the sample on the last seven items. The second explanation is that this respondent is actually of very high ability (or whatever construct is supposed to be measured with these items) but was at a disadvantage on the items in the middle of this row (perhaps because of coding alignment

errors, misinterpretation of items, etc.). Either way, there is an interaction between the respondent and some characteristic of the items. The problem is that we have no way of knowing which is the correct explanation. In other words, we have no way of knowing the true standing of this respondent on the construct of interest, i.e., the test is inappropriate as a measure of the construct of interest.

It is important to note that, if the advantage or disadvantage of the respondent does not produce inconsistency (i.e., does not force a departure from a Guttman vector), then none of these indices (IRT-based or otherwise, standardized or otherwise) will detect it. However, if no inconsistency is produced, then there cannot be a person by item interaction. Instead, there would be a simple main effect for the personal characteristic, and the inappropriateness would be Type P inappropriateness and not Type I inappropriateness.

Sato's Caution index and others like it are designed to detect just this sort of interaction. The main problem with these indices is that they are highly related to total score. Specifically, respondents with very high or very low total scores are more likely to be identified as aberrant because there is more room for aberrance. For example, a respondent with a very low total score who happens to answer one or two of the more difficult items correctly will likely receive a large score on any index that is not well-standardized simply because those one or two item responses are so inconsistent with the total score of the respondent whereas the same situation applied to a respondent with an average score will produce an index value that is not nearly as extreme. Since the goal of inappropriateness measurement is to measure inappropriateness independent of total score (or theta), this is seen as a disadvantage to poorly standardized measures.

IRT-based indices of Type I inappropriateness (unstandardized). There are many IRT-based indices of Type I inappropriateness, and they are usually based either on the Rasch model (one-parameter, 1960) or the three-parameter model (Hambleton & Cook, 1977). These indices address the question: To what extent do the responses of a given respondent conform to the Item Characteristic Curves of the items in a test? In terms of Type I inappropriateness: To what extent is there something about this respondent that renders the Item Characteristic Curves invalid for this respondent? In other words, To what extent is there an interaction between the personal characteristics of the respondent and characteristics of the items? This question also applies to both Maximum and typical tests. In the Rasch model approaches, the ICC's differ only with respect to the difficulty parameter, whereas in the three-paremeter model approaches, the ICC's differ with respect to difficulty, discrimination, and the pseudo-guessing parameter.

An example of an index based on the Rasch model is the unweighted total fit mean square (U1) discussed in Wright and Panchapakesan (1969). The formula for U1 is

$$UI_i = \frac{\sum_{j=1}^{N}\left[\frac{(u_{ij}-P_{ij})^2}{P_{ij}(1-P_{ij})}\right]}{N}$$

where i indexes the examinee, j indexes the N items, P_{ij} is the probability of a correct response predicted by the Rasch model, and u_{ij} is the observed item response.

This is essentially a measure of the average discrepancy between the observed responses of a given examinee and the responses predicted by the model. The larger the discrepancy is, the more caution should be used in interpreting the total score, i.e., the greater the degree of inappropriateness.

An example of an index of Type I inappropriateness based on the 3-parameter model is the l_0 index described by Levine and Rubin (1979). The formula for l_0 is

$$L(\theta_i = \prod_{j=1}^{N} P_{ij}^{u_{ij}} (1 - P_{ij})^{(1-u_{ij})}$$

where P_{ij} is the probability of a correct response based on the three parameter model and u_{ij} is the observed response.

This is the log of the compound probability of the observed response pattern for a maximum likelihood estimate of ability (Rudner, 1983). The rationale for this index is similar to that of the U1 index: l_0 is a measure of the discrepancy between the observed responses and the responses predicted by an IRT-model, specifically, the three-parameter model. l_0 is perhaps the most widely cited index of Type I inappropriateness.

The problem with these two indices, as with the non-IRT based indices discussed above, is that they are poorly standardized, that is, they are highly related with total score (Rudner, 1983; Birenbaum, 1985). The solution is, of course, to attempt to develop indices that are well-standardized and, therefore, relatively unrelated to total score. The next two sections are devoted to just such indices.

Standardized, non-IRT based indices of Type I inappropriateness. One example of a standardized, non-IRT based index is the personal biserial of Donlon and Fischer (1968). Although this index has been shown to be related to total score, it is useful as an illustration of the meaning of standardization. The personal biserial coefficient is simply the biserial correlation between the dichotomously scored item responses of a given respondent and the difficulties of those items. The reason I claim that this index is standardized is that it is the biserial correlation. Since the biserial correlation is insensitive to the variances of the variables being correlated, the total score of a given respondent, which is based on the proportion of correct responses given by the respondent, should have little effect on the personal biserial (as opposed to the point biserial correlation between responses and difficulties).

As was mentioned above, the personal biserial has been found to be highly related to total score, which means that it is not well standardized. A better example of a standardized, non-IRT based index is the Modified Caution Index (MCI, Harnisch & Linn, 1981).

$$C_i^* = \frac{\sum_{j=1}^{n_i}(1-\mu_{ij})n_j - \sum_{j=n_i+1}^{J}\mu_{ij}n_j}{\sum_{j=1}^{n_i}n_j - \sum_{j=J+1-n_i}^{J}n_j}$$

where the symbols are the same as those in Sato's original index (Equation 1)

This is simply Sato's Caution Index (Sato, 1975) modified to yield a lower bound of 0 and an upper bound of 1, thus eliminating the extreme scores that can be obtained on the caution index for very high scoring examinees who miss a single very easy item or for very low scoring examinees who answer correctly a single very difficult item. The MCI index has been found to have little or no relationship with total score (Harnisch & Linn, 1981; Rudner, 1983). It is, therefore, considered to be a well-standardized index.

Standardized, IRT-based indices of Type I inappropriateness.

These are by far the most commonly used and studied statistical indices of appropriateness and, therefore, the most common indices of Type I inappropriateness. Two examples of such indices are the standardized extended caution index of Tatsuoka and Tatsuoka (1982) and the standardized l_0 index (l_z) of Drasgow, Levine, and Williams (1985). Since these two have been found to be highly correlated (Birenbaum, 1985), and since l_z has been applied to a wider variety of situations (e.g., Drasgow, Levine, & McLaughlin, 1991; Drasgow, Levine, McLaughlin, Williams, & Candell, 1989), the l_z index will be the focus of the present paper.

The formula for l_z is

$$l_z = \frac{l_0 - E(l_0)}{\sqrt{[Var(l_0)]}}$$

where

$$E(l_0)=\sum_{i=1}^{n}\left[P_i(\hat{\theta})\ln P_i(\hat{\theta})+[1-P_i(\hat{\theta})]\ln[1-P_i(\hat{\theta})]\right]$$

and

$$Var(l_0)=\sum_{i=1}^{n}P_i(\hat{\theta})[1-P_i(\hat{\theta})]\left[\ln\frac{P_i(\hat{\theta})}{1-P_i(\hat{\theta})}\right]^2$$

l_z is approximately normally distributed with a mean of 0 and a standard deviation of 1. A negative l_z indicates inconsistency of the pattern of responses. The literature on l_z has focussed exclusively on the negative form of the index, with an l_z value of -1.65 (i.e., the score from the standard normal distribution corresponding to a level of significance of .05 for a one-tailed test) indicating an aberrant response pattern. A positive l_z indicates hyperconsistency, or a pattern of responses that fits the IRT model so well that it is suspicious. Positive l_z's have received virtually no attention in the appropriateness literature, and their meaning is largely unclear.

The l_z index is a measure of goodness of fit of an IRT model to a particular response pattern. In other words, l_z measures the extent to which a given response pattern is determined by factors other than ability (or the noncognitive equivalent) and the parameters of the three-paremeter model.

The study of appropriateness measurement with l_z has been almost completely statistical. There has been virtually no attempt to assess the construct validity of l_z or any

of the measures of Type I inappropriateness. It is known that l_z detects departure from the IRT model, and there have been numerous suggestions as to the possible causes of such a departure (e.g., cheating, coding errors, test anxiety), but no attempt has been made to model these causes as determinants of l_z. We know only that these indices measure the consistency of response patterns relative to item characteristics. As Reise (1990) points out, however, these indices tell us of response inconsistency, not why the responses are inconsistent. As a result, we have no clear idea of what l_z and other similar indices are measuring. I maintain that these indices are reflective of person by item interactions and that viewing inappropriateness in this way will lead to a greater understanding of the construct that one intends to measure as well as the nature of inappropriateness.

I have offered various sources of Type I inappropriateness and rationale for their effects on item responses vis a vis appropriateness, and suggest further that it is precisely these sources that are captured by l_z. In this way, I offer an assessment of the construct validity of l_z. Specifically, I suggest the following extensions of my earlier propositions:

> Proposition 1A - 8A - The l_z index becomes more extreme as the interaction between characteristics of the respondent (e.g., test anxiety, test wiseness, etc.) and characteristics of the items (e.g., ambiguity of meaning, complexity of response options, etc.) becomes more pronounced. Specifically, for all of the interactions involving respondent characteristics other than omissiveness, l_z becomes more extreme in the negative direction as the interaction effect becomes stronger. For the interactions that do

involve omissiveness, l_z becomes more extreme in the positive direction as the interaction becomes stronger.

The interactions involving omissiveness are expected to produce positive l_z values because they are expected to produce hyperconsistency. Those respondents high in omissiveness are expected to omit those items that are too difficult for them, which means that they have no chance of answering them correctly, which in turn would produce a near-perfect Guttman vector (i.e., a hyperconsistent response pattern).

To the extent that this set of propositions is borne out, the construct validity of the l_z index and other indices like it will be more firmly established. Also relevant to the issue of construct validity is the fact that l_z is not hypothesized to detect any of the sources of Type P inappropriateness. Since the sources of Type P inappropriateness alone do not lead to inconsistency of response patterns, they should not be detected by l_z except insofar as they are related to their respective interactions. This suggests a final model of appropriateness as measured by l_z.

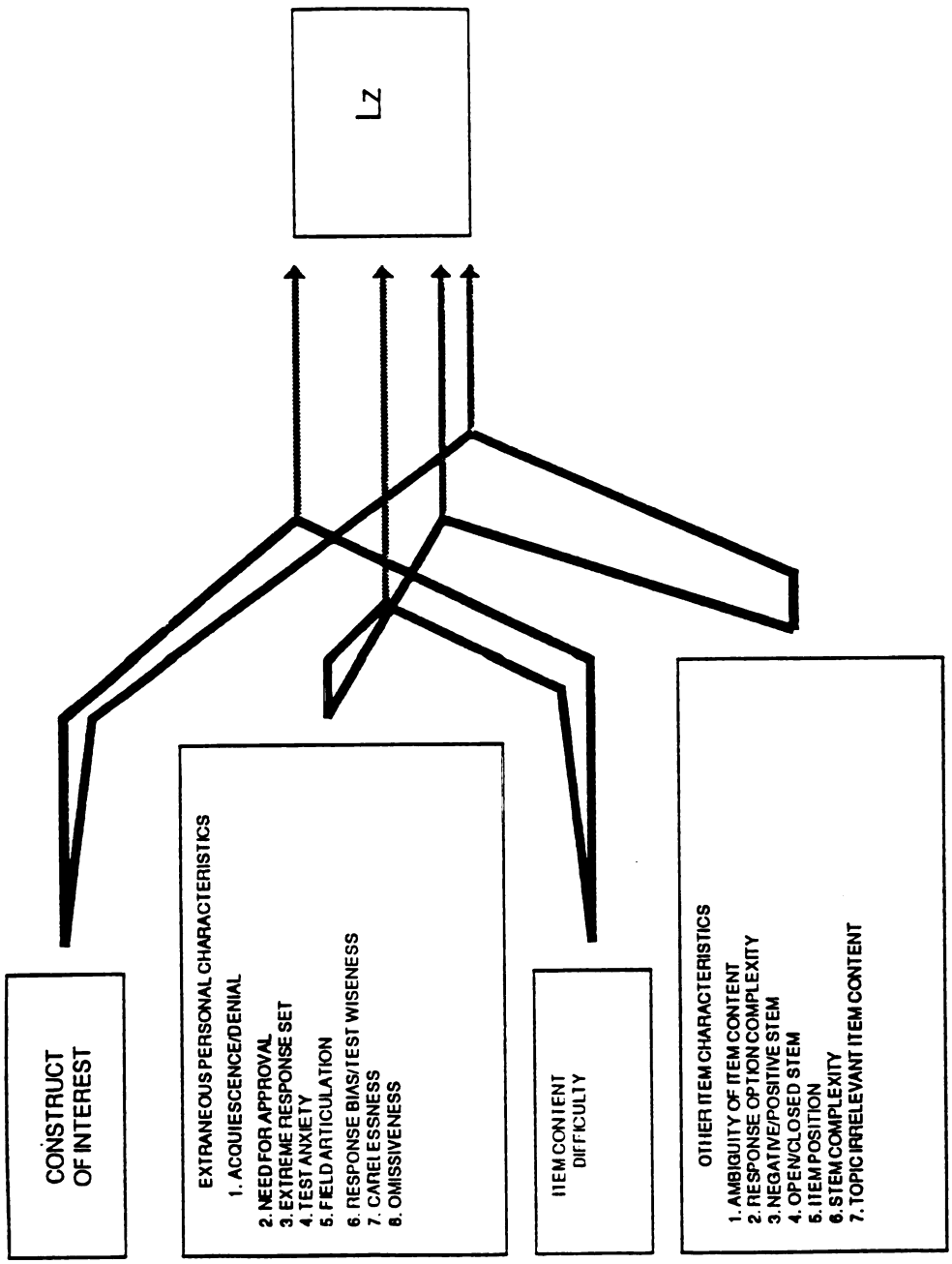

Figure 10. A model of the determinants of l_z

Overall summary

Sources of test inappropriateness and research on these sources were discussed. Two types of sources were identified: those involving main effects for respondent characteristics on item responses (Type P) and those involving interactions between respondent characteristics and item characteristics (Type I). Measures of these sources were also discussed. Sources of inappropriateness were discussed in terms of both Maximum and typical tests. Table 1 summarizes the sources of inappropriateness that were discussed.

Table 1
Sources of Type P and Type I inappropriateness for both Maximum and typical tests

SOURCE	MAXIMAL		typical	
	Type P (Personal characteristics)	TYPE I (Interaction)	Type P (Personal characteristics)	T Y P E I (Interaction)
Acquiescence/ Denial	excessive selection of one or the other of true/false options		excessive agreement/ denial	
Need for Approval	cheating	nApp X opportunity for cheating	faking good	nApp X social des. of items
Extreme response set,bias/central tendency	excessive selection of options in particular positions	see response bias	excessive selection of options in particular positions	response set X suscept. of items to response set
Test Anxiety	difficulty processing test items	test anxiety X difficulty-based item characteristics	difficulty processing items	test anxiety X difficulty-based item characteristics
Field Articulation	distinction between relevant and irrelevant item content	field articulation X distracting characteristics of items	distinction between relevant and irrelevant item content	field artic. X distracting item characteristics

Table 1 cont'd

SOURCE (Interaction)	MAXIMUM Type P (Personal characteristics)	MAXIMUM TYPE I (Interaction)	typical Type P (Personal characteristics)	typical T Y P E I
Response bias/ Test Wiseness	selection of responses based on option position or wiseness	bias/wiseness X susceptibility of items to these factors	selection of responses based on wiseness	wiseness X susceptibility of items to wiseness
Carelessness/ Motivation	careless responding	motivation X difficulty-based item characteristics	careless responding	motivation X difficulty-based item characs.
Omissiveness	lack of response to difficult items	omissiveness X difficulty-based item characteristics	lack of response to difficult items	omissiveness X difficulty-based item characs.

The present study

The purpose of the present study was to test a part of this model. Specifically, I examined the effects on knowledge test scores and test inappropriateness as measured by l_z of test anxiety, math anxiety (which can be viewed as a specific application of test anxiety. See p.15 for the reference to "number anxiety"), conscientiousness, carelessness, difficulty-based item characteristics such as positive negative wording, open/closed stem format, and response option complexity, and the interaction between each of the four respondent characteristics mentioned above and difficulty-based item characteristics. The details of the present study are described below. The specific hypotheses tested in this study are as follows.

Hypothesis 1 - Difficulty-based item characteristics have a deleterious effect on knowledge test performance.

Hypothesis 2 - Respondents who are higher in conscientiousness have higher knowledge test scores than do respondents who are lower in conscientiousness.

Hypothesis 3 - Respondents who are higher in carelessness have lower knowledge test scores than do respondents who are lower in carelessness.

Hypothesis 4 - Respondents who are higher in math anxiety have lower knowledge test scores than do respondents who are lower in math anxiety

Hypothesis 5 - Respondents who are higher in test anxiety have lower knowledge test scores than do respondents who are lower in test anxiety

Hypothesis 6 - The effect of conscientiousness on knowledge test scores is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that the knowledge test responses of those

respondents higher in conscientiousness are less adversely affected by difficulty-based item characteristics than are those of respondents lower in conscientiousness.

Hypothesis 7 - The effect of carelessness on knowledge test scores is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that the knowledge test responses of those respondents higher in carelessness are more adversely affected by difficulty-based item characteristics than are those of respondents lower in carelessness.

Hypothesis 8 - The effect of math anxiety on knowledge test scores is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that the knowledge test responses of those respondents higher in math anxiety are more adversely affected by difficulty-based item characteristics than are those of respondents lower in math anxiety.

Hypothesis 9 - The effect of test anxiety on knowledge test scores is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that the knowledge test responses of those respondents higher in test anxiety are more adversely affected by difficulty-based item characteristics than are those of respondents lower in test anxiety.

Hypothesis 10 - The effect of conscientiousness on l_z values is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that the l_z values will be negatively related to the presence of difficulty-based item characteristics in test items for those respondents lower in conscientiousness but relatively unrelated for those of respondents higher in

conscientiousness. In other words, difficulty-based item characteristics will lead to inappropriateness for those respondents low in conscientiousness but not for those respondents high in conscientiousness.

Hypothesis 11 - The effect of carelessness on l_z values is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that l_z values will be negatively related to the extent to which difficulty-based item characteristics are present in test items for those respondents higher in carelessness but relatively unrelated for those of respondents lower in carelessness. In other words, difficulty-based item characteristics will lead to inappropriateness for those respondents high in carelessness but not for those respondents low in carelessness.

Hypothesis 12 - The effect of math anxiety on l_z values is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that l_z values will be negatively related to the extent to which difficulty-based item characteristics are present in test items for those respondents higher in math anxiety but relatively unrelated for those of respondents lower in math anxiety. In other words, difficulty-based item characteristics will lead to inappropriateness for those respondents high in math anxiety but not for those respondents low in math anxiety.

Hypothesis 13 - The effect of test anxiety on l_z values is moderated by the extent to which the knowledge test items contain difficulty-based item characteristics such that l_z values will be negatively related to the extent to which difficulty-based item characteristics are present in test items for those respondents higher in test

anxiety but relatively unrelated for those of respondents lower in test anxiety. In other words, difficulty-based item characteristics will lead to inappropriateness for those respondents high in test anxiety but not for those respondents low in test anxiety.

Method

Sample

Subjects were 165 undergraduates from a large, midwestern university. 67% of the subjects were women. No other demographic data were collected. They were recruited from Introductory Statistics classes towards the end of the semester so that they had had an opportunity to learn most of the course material. Subjects were given extra credit for participation. This sample, while convenient, was also quite appropriate for the variables examined in this study. The general focus of this study was testing, with emphasis on the relationships among respondent characteristics, item characteristics, and construct validity of items. Since testing is a common part of most university educations, a sample of college students was ideal for the examination of factors that affect testing.

Design

The present study used a repeated measures regression design with four between subjects factors, one within subjects factor, and two dependent variables. The between-subjects predictor variables were Conscientiousness, Math Anxiety, Test Anxiety, and Carelessness. The within-subjects factor was Item Difficulty as determined by item construction principles. The dependent variables were scores on tests of statistical knowledge and the consistency of responses to items from those tests. Thus, eight

repeated measures regression analyses were performed, one for each combination of between-subjects variable and dependent variable.

Measures

Conscientiousness.

Conscientiousness was assessed with four items from the twelve-item conscientiousness scale of the NEO-PI personality inventory (Costa & McCrae, 1991). These four items were the items in the scale which related directly to dependability (as opposed to organizational skills and goal orientation; see Appendix A). The Costa & McCrae (1991) measure was used because it is one of the few questionnaire measures designed specifically to assess conscientiousness as defined by proponents of the Big Five theory of personality (e.g., Digman, 1990). Internal consistency reliability for the four items was estimated to be .68, suggesting that uniquenesses for the four items were acceptable (Cortina, 1993).

Math Anxiety.

Math Anxiety was assessed with the 11-item Math Anxiety Questionnaire developed by Wigfield & Meece (1988), which in turn was based on a measure which was originally developed by Richardson & Suinan (1972). This measure was used because it taps both the emotional and cognitive components of anxiety. Internal consistency reliability for the four items was estimated to be .85, suggesting that uniquenesses for these items were also acceptable (Cortina, 1993).

Test Anxiety.

Test Anxiety was assessed with the 10 - item Test Anxiety Scale developed by Morris, Davis, & Hutchings (1981). This measure improves upon earlier test anxiety

scales (e.g., Mandler & Sarason, 1952) which failed to tap both the emotional and cognitive components of anxiety. Internal consistency reliability for the four items was estimated to be .83, suggesting that uniquenesses for these items were also acceptable (Cortina, 1993).

Carelessness.

Carelessness was assessed with a six-item scale constructed by the author. These items were similar to the "nonsense" items included in many noncognitive tests in that they were designed to produce a particular response from any respondent who pays attention to (i.e., reads) the item. The unique aspect of the items that make up the carelessness scale used in the present study is that they are not easily recognized as items which tap carelessness. Typical carelessness items are absurdities which can be recognized by respondents who are merely scanning items. Such identification can have a deleterious influence on test taking motivation. The items used in the present study were statistical knowledge items that were answered correctly by all respondents during all pretesting situations (details are described below). Items such as

Another word for the average is

a) mean

b) variance

c) standard deviation

d) range

should be answered correctly by any Introductory Statistics student who reads the question, as was the case during pretesting. Any variability in responses to such items should be due only to carelessness.

Statistics knowledge test.

All subjects were administered the 75-item test of statistical knowledge contained in Appendix B. The items on the statistical knowledge test were items typically found on exams for Introductory Statistics classes. Items with three levels of content-irrelevant difficulty were developed. 24 items were open-stemmed, negatively worded, contained complex response options, and had complex stems (e.g., word problems), using the definition of stem complexity from Zimmerman (1954). These were the "Difficult" items. The following is an example of one of the "difficult" items:

Difficult item - Suppose I know the number of times each Michigan resident has been swindled by Gov. Engler (So, I have access to this population of scores). I then take many different samples of 15 people each and calculate the mean for each sample. If the mean of the means were 7.8 and the standard deviation of individual scores were 2.2, the population mean and the standard error of the mean would not be

a. mu = 2.79, sigma = .57

b. mu = 7.8, sigma = 2.2

c. either a or b

d. all of the above

25 Moderately difficult items were similar to the Difficult items except that they were positively worded and closed stemmed. The following is an example of one of the "moderate" items:

Moderate item - For some strange and terrible reason, I am interested in knowing the average number of white collar crimes committed per day by Sen. Bob Dole over the past 2000 days. In an attempt to estimate this value, I randomly choose twenty days from these 2000, count the number of white collar crimes he committed on each of those 20 days, and get the average of those twenty numbers.

What is the statistic that I have used?

a. The average number of white collar crimes per day committed by Dole over the past 2000 days.

b. The average number of white collar crimes committed per day by Dole over the 20 days that I measured.

c. 2000

d. all of the above

26 Easy items were similar to the Moderately difficult items except that the stems were noncomplex and there were no complex response options. The following is an example of one of the "easy" items:

Easy item - Which of the following is an advantage of the mean as a measure of central tendency?

a. It is greatly affected by extreme scores

b. It can be manipulated algebraically

c. It is not greatly affected by extreme scores

d. It is difficult to calculate

Participants' scores on the items within each level of difficulty were collapsed to form single variables for the regression analyses involving knowledge test scores described below. Difficulty as defined in this paragraph refers to aspects of the item format that are thought to decrease the proportion of correct responses independently of the examinees' knowledge of the domain being assessed.

An attempt was made to equate items with respect to content difficulty (i.e. construct relevant difficulty). The reason for this is that the item characteristics that are of the most concern to test constructors are those over which they have direct control. While most tests should and do vary with respect to content difficulty, other item characteristics such as option complexity and stem complexity can and should be controlled, especially if they foster aberrant response patterns. Items for the statistical knowledge test were chosen from a pool of test items that had been administered to undergraduates as items in actual tests. Specifically, 75 items which contained none of the difficulty-inducing item characteristics mentioned earlier (e.g., complex response options, negative wording, etc.) were chosen and distributed randomly into one of the three groups. Inspection of the item difficulty values (percentage incorrect) calculated from these previous testing situations showed that the average item difficulties (proportion answering the item incorrectly) within the three groups were almost identical (.26 for items which were to be used for the "easy" test, .23 for items which were to be used for

the "moderate" test, and .25 for the items which were to be used for the "difficult" test), suggesting that these three groups of items were virtually identical with respect to content-relevant difficulty. Differences in test scores across difficulty levels as defined above can then be attributed to the manipulations of item characteristics. The measure of statistical knowledge used was proportion of correct responses.

Response Consistency - Response consistency was assessed with the l_z index of Drasgow et al. (1985). The l_z index requires the calculation of the item parameters of the three-parameter IRT model. These parameters were calculated from the responses of subjects to the 75 test items.

Procedure

Subjects were first approached during their statistics classes and asked if they would be willing to participate in the experiment. Those who agreed were asked to sign up for a testing date as well.

The 75-item statistics knowledge test and the four tests measuring respondent characteristics were administered to large groups of subjects at a time. The measures of conscientiousness, math anxiety, and test anxiety were administered first, followed by the knowledge test. Two forms of the test were created. The two forms differed only in that the order in which the items were presented was reversed. Within each form, item order was random. The purpose of generating two forms was to allow an examination of order effects. Neither knowledge test scores nor l_z values differed across the two forms.

The items measuring carelessness were embedded within the knowledge test. Since all 75-items were administered to all subjects, all three levels of difficulty were experienced by all subjects. In an attempt to increase motivation to respond carefully, the

test administrators explained to subjects that the test results would be used by the Instructor to evaluate his own teaching performance. They were also told that the test provided an opportunity to practice for upcoming tests in the class. Finally, $100 was awarded to each of three of the top performers on the knowledge test.

Data Analysis

After establishing the unidimensionality of the knowledge test, the responses of subjects to the test were analyzed with the BILOG IRT computer program (Mislevy & Bock, 1990). This analysis yields both ability estimates for respondents (Θ) and item parameter estimates that are necessary to compute l_z as outlined above in the introduction.

Hypotheses were tested with repeated measures hierarchical regression (RMHR: Cohen & Cohen, 1983; Hollenbeck, Ilgen, & Sego, in press). As there was no a priori rationale for investigating the effects of the four between-subjects predictors in conjunction with one another, separate regression analyses were performed for each of the four between-subjects variables (conscientiousness, math anxiety, test anxiety, and carelessness) and each of the two criteria (percentage of knowledge items answered correctly and l_z) for a total of eight regressions. In each regression, the dependent variable (knowledge test scores or l_z) was regressed onto one of the between-subjects factors and the within-subjects factor. The details of this procedure are described below. It was expected that the interaction between item characteristics and each of the four between-subjects factors would explain a significant portion of the relevant variance in knowledge test scores above and beyond that explained by the main effects for the predictor variables, and that insofar as these interactions were significant, they would also explain relevant variance in l_z.

Results

Tests measuring respondent charactersitics

Table 2 contains means, standard deviations, intercorrelations, and internal consistency estimates for the Conscientiousness, Math Anxiety, Test Anxiety, and Carelessness Scales. Means, standard deviations, and intercorrelations are presented for l_z. The values presented for the knowledge tests refer to the knowledge tests composed of the items that remained after the initial BILOG analysis (see below for details of this analysis).

Table 2

Descriptive statistics for all tests and l$_z$'s

	Mean	SD	α	1	2	3	4	5
1. Conscien.	15.86	2.39	.68					
2. Math Anx.	35.33	8.16	.85	-.05				
3. Test Anx.	18.51	5.94	.83	-.16*	.48*			
4. Careless.	5.32	.87	.28	.01	-.06	-.04		
5. Easy Test	.57	.16	.69	.01	-.21*	-.12	.37*	
6. Mod. Test	.54	.17	.70	.11	-.22*	-.10	.36*	.57*
7. Diff. Test	.46	.16	.64	-.06	-.17*	-.08	.44*	.51*
8. lz (easy)	.09	1.01		-.07	.08	-.02	-.15	-.10
9. lz (mod)	.05	.92		.04	-.03	-.05	-.14	.04
10. lz (diff)	-.06	1.02		.08	.08	.03	-.05	-.06

Table 2 cont'd

	6	7	8	9
7. Diff. Test	.55*			
8. lz (easy)	-.26*	-.01		
9. lz (mod)	.04	-.05	-.07	
10. lz (diff)	.03	-.17*	-.14	.10

* - $p<.05$

There are several points to be made with respect to this table. Regarding the measures of respondent characteristics, there was reasonable variability on the Conscientiousness, Math Anxiety, and Test Anxiety scales. Also, the means for these scales were comparable to those reported in previous literature (e.g., Morris et al., 1981; Wigfield & Meece, 1988; Costa & McCrae, 1988). There was considerably less variability in the Carelessness measure, but this is not surprising given the simple nature of the questions in the scale.

Internal consistency estimates for the conscientiousness, math anxiety, and test anxiety scales were adequate suggesting acceptable levels of item uniqueness (Cortina, 1993). Although the estimate for the conscientiousness scale in the present study was lower than those presented in previous research, this is not surprising given the fact that the estimate in the present study was based on only four items.

Internal consistency for the Carelessness scale, however, was quite low (α=.28). Again, this is not surprising given the fact that a substantial portion of respondents answered all of these items in the same way.

Table 2 also contains information about the statistical knowledge items and l_z values. This information is discussed below.

Statistical knowledge test

The "easy", "moderate", and "difficult" tests were composed of 26, 25, and 24, items respectively. Item means, standard deviations, and intercorrelations can be found in Appendix C.

Before IRT analyses were performed, the dimensionality of the items was assessed with a factor analysis of the interitem correlation matrix after that matrix was transformed

into a matrix of polychoric correlations. Table 3 contains the eigenvalues and percentage of variance accounted for all factors with eigenvalues greater than 1.

Table 3

Factor analysis of knowledge test items

FACTOR	EIGENVALUE	PCT OF VAR
1	11.10568	14.8
2	4.06301	5.4
3	3.88617	5.2
4	3.41616	4.6
5	3.14234	4.2
6	3.04511	4.1
7	2.92488	3.9
8	2.78527	3.7
9	2.65622	3.5
10	2.49907	3.3
11	2.38604	3.2
12	2.20712	2.9
13	2.16898	2.9
14	2.03637	2.7
15	1.97130	2.6
16	1.89015	2.5
17	1.75348	2.3
18	1.65575	2.2
19	1.61516	2.2
20	1.57926	2.1
21	1.52589	2.0
22	1.45656	1.9
23	1.36599	1.8
24	1.30923	1.7
25	1.25342	1.7
26	1.19881	1.6
27	1.16056	1.5
28	1.08787	1.5
29	1.01588	1.4

The conclusion to be drawn with respect to dimensionality depends on the criterion that one uses. There are many factors with eigenvalues greater than 1. Given the range of knowledge tapped by the test and the range of item characteristics, this is not surprising. However, only one of the factors explains more than 5.4% of the test variance. Also, the first factor eigenvalue is almost three times the size of the next largest eigenvalue (11.11 vs. 4.06; Hulin et al., 1983). Given the latter two facts, IRT analysis was deemed appropriate.

Item parameter estimates for the 75 items and θ-parameter estimates for the 165 respondents were generated with BILOG (Mislevy & Bock, 1979). This analysis suggested that five of the items (Nos. 25, 41, 46, 64, and 75) did not conform to the three-parameter IRT model. X^2 values and degrees of freedom for these items were 8436.2 (<1df), 5.1 (4df), 7480.8 (<1df), 9.3 (5df), and 13.5 (4df). These items were then discarded, leaving 26 "easy" items, 23 "moderate" items, and 21 "difficult" items to be reanalyzed with BILOG. Because there were different numbers of items in the three tests, subsequent analyses involved test means instead of simple raw score composites.

Item parameter estimates from this second BILOG analysis as well as corresponding X^2 values can be found in Appendix D. As expected, discrimination and guessing parameter estimates were similar across the three types of test items while difficulty parameter estimates were considerably higher for the "difficult" items than for the "easy" and "moderate" items (mean difficulty parameter estimates were .91 for the "difficult" items as opposed to .42 and .38 for the "easy" and "moderate" items).

These three sets of items were then combined to form three knowledge scales. As was mentioned above, mean scores across items for these tests were used instead of simple sums to control for the different numbers of items in each knowledge scale. Means, standard deviations, internal consistency estimates, and intercorrelations for these three tests are presented in Table 2. The differences in means suggest that the item characteristics manipulated in this study affected test scores in the manner predicted, i.e., the items with the more difficult formats were answered incorrectly more often than were the items with the less difficult formats.

Test scores were significantly correlated with math anxiety and carelessness: those respondents who were higher in math anxiety received lower scores on the knowledge tests, and those respondents who answered more of the carelessness items correctly received higher scores on the knowledge tests (as expected). The correlations involving carelessness are particularly compelling given the relative lack of variability on the carelessness measure.

Internal consistency estimates for the knowledge tests were marginal, but this is not surprising given the range of content within each of the three tests.

l_z

The item and theta parameter estimates generated by the BILOG program were used to compute l_z values for each respondent for each test. Means and standard deviations for these l_z's can be found in Table 2. As can be seen, there are trivial mean differences in l_z values across the three tests. As expected, l_z values, which are standardized with respect to theta, had little or no relationship with knowledge test scores. l_z's were, however, unrelated to any of the respondent characteristics,

suggesting that there is little or no main effect for these measures on appropriateness as measured by l_z.

Tests of Hypotheses

Hypotheses were tested using repeated measures regression (Cohen & Cohen, 1983; Gully, 1994). This procedure involves several steps. First, variance of the criterion variable is partitioned into that attributable to between-subject differences and that attributable to within-subject differences. In the case of the present study there were two such criterion variables: percentage of correct responses and l_z values. Table 4 contains variance attributable to between- and within-subjects effects as well as percentages for each of these values for each of the two criteria.

Table 4

Variance of dependent variables attributable to between- and within-subjects effects

Dependent Variable	Btwn. Ss Variance	W/in Ss Variance	% of s^2 for Btwn.	% of s^2 for W/in
Knowledge Test Scores	.0196	.0093	67.8	32.2
l_z	.2916	.6684	30.4	69.6

The table shows that approximately two-thirds of the variance in knowledge test scores was between-subjects variance while approximately two-thirds of the variance in l_z was within-subjects variance. This suggests that between-subjects predictors are more likely to explain the variability in test scores whereas within-subjects predictors are more likely to explain the variability in l_z.

The second step in the procedure involved the regression of the variance attributable to within-subjects differences onto all within-subjects predictors and interactions involving only within-subject predictors. In all of the analyses for the present study, there were two such within-subjects predictors. These two predictors corresponded to the two dummy codes created to represent the three levels of difficulty-based item characteristics. The F-tests for the within-subjects effects were performed on P_{w1}, #obs-N-P_{t1} degrees of freedom where P_{w1} is the number of within-subjects predictors entered in the first step, P_{t1} is the total number of within-subjects predictors entered as of that step, #obs equals the number of total observations, and N equals the total sample size. In those analyses involving percentage of correct responses on the knowledge test, this first step in the regressions was performed with 2 and 330 degrees of freedom. In those analyses involving l_z, this first step in the regressions was performed with 2 and 328 degrees of freedom.

In the third step of the procedure, the variance attributable to between-subjects differences is regressed on all between-subjects predictors and interactions involving only between-subjects predictors. In all of the analyses for this study, there was one such main effect and no between-subjects interactions. The degrees of freedom for this step are P_b, N-P_b where P_b is the number of between-subjects predictors entered in

this step. In those analyses involving percentage of correct responses on the knowledge test, this second step in the regressions was performed with 1 and 164 degrees of freedom. In those analyses involving l_z, this second step in the regressions was performed with 1 and 163 degrees of freedom.

Finally, remaining within-subjects variance is regressed onto interactions involving both within- and between-subjects predictors. In all analyses for this study, there were two such interactions: one for each of the two dummy variables. The formula for the degrees of freedom for this step is P_{w2}, #obs-N-P_{t2} where P_{w2} is the number of predictors entered at this step and P_{t2} is the total number of predictors entered as of this step. In those analyses involving percentage of correct responses on the knowledge test, this final step in the regressions was performed with 2 and 328 degrees of freedom. In those analyses involving l_z, this second step in the regressions was performed with 2 and 326 degrees of freedom.

All of the hypotheses in this study were tested with this procedure. As was mentioned above, separate regression analyses were conducted for each of the four between-subjects predictors in question and each of the two criteria. Although the fact that the four between-subjects predictors were correlated with one another means that the regressions were somewhat redundant, the separate analysis of these predictors eliminated the need to tease apart the effects of predictors analyzed in conjunction. Table 5 contains the correlations among all terms used in the regression analyses that are described below.

Table 5

Intercorrelations among all terms used in regression analyses

TERM	CONS	CARE	MANX	TANX	DMOD[1]
CONS	1.0000	.0126	-.0558	-.1650*	.0000
CARE	.0126	1.0000	-.0641	-.0457	.0000
MANX	-.0558	-.0641	1.0000	.4827*	.0000
TANX	-.1650*	-.0457	.4827*	1.0000	.0000
DMOD	.0000	.0000	.0000	.0000	1.0000
DDIF	.0000	.0000	.0000	.0000	.5000
MCON	.1045*	.0013	-.0058	-.0172	.9835*
DCON	.1045*	.0013	-.0058	-.0172	-.4917*
MCAR	.1045*	.0013	-.0058	-.0172	.9835*
DCAR	.1045*	.0013	-.0058	-.0172	-.4917*
MMAT	-.0087	-.0100	.1567*	.0757	.9624*
DMAT	-.0087	-.0100	.1567*	.0757	-.4812*
MDANX	-.0087	-.0100	.1567*	.0757	.9624*
DFANX	-.0087	-.0100	.1567*	.0757	-.4812*
LZ	.0173	-.1141*	.0457	-.0128	-.0131
KWTST	.0206	.3790*	-.1968*	-.0981*	.0766

Table 5 cont'd

DDIF	MCON	DCON	MCAR	DCAR	MMAT
CONS .0000	.1045*	.1045*	.1045*	.1045*	-.009
CARE .0000	.0013	.0013	.0013	.0013	-.010
MANX.0000	-.0058	-.0058	-.0058	-.0058	.157*
TANX .0000	-.0172	-.0172	-.0172	-.0172	.077
DMOD-.5000*	.9835*	-.4917*	.9835*	-.4917*	.962*
DDIF 1.0000	-.4917*	.9835*	-.4917*	.9835*	-.481*
MCON-.4917*	1.0000	-.4836*	1.0000	-.4836*	.944*
DCON .9835*	-.4836*	1.0000	-.4836*	1.0000	-.473*
MCAR-.4917*	1.0000	-.4836*	1.0000	.4836*	.945*
DCAR .9835*	-.4836*	1.0000	.4836*	1.0000	-.473*
MMAT-.4812*	.9438*	-.4733*	.9438*	-.4733*	1.000
DMAT .9624*	-.4733*	.9438*	-.4733*	.9438*	-.463*
MDANX-.4812*	.9438*	-.4733*	.9438*	-.4733*	1.000
DFANX.9624*	-.4733*	.9438*	-.4733*	.9438*	-.463*
LZ .0050	-.0088	.0038	-.0088	.0038	-.019
KNTST-.2658*	.0870	-.2508*	.0870	-.2508*	.038

Table 5 cont'd

	DMAT	MDANX	DFANX	LZ	KWTST
CONS	-.0087	-.0087	-.0087	.0173	.0206
CARE	-.0100	-.0100	-.0100	-.1141*	.3790*
MANX	.1567*	.1567*	.1567*	.0457	-.1968*
TANX	.0757	.0757	.0757	-.0128	-.0981*
DMOD	-.4812*	.9624*	-.4812*	-.0131	.0766
DDIF	.9624*	-.4812*	.9624*	-.0050	-.2489*
MCON	-.4733*	.9438*	-.4733*	-.0088	.0870
DCON	.9438*	-.4733*	.9438*	.0038	-.2508*
MCAR	-.4733*	.9438*	-.4733*	-.0088	.0870
DCAR	.9438*	-.4733*	.9438*	.0038	-.2508*
MMAT	-.4631*	1.0000	-.4631*	-.0176	.0380
DMAT	1.0000	-.4631*	1.0000	.0087	-.2649*
MDANX	-.4631*	1.0000	-.4631*	-.0176	.0380
DFANX	1.0000	-.4631*	1.0000	.0087	-.2649*
LZ	.0087	-.0176	.0087	1.0000	-.0688
KWTST	-.2649*	.0380	-.2649*	-.0688	1.0000

[1] - CONS=Conscientiousness;CARE=Carefulness;MANX=Math Anxiety;TANX=Test Anxiety;DMOD=Dummy for "moderate" test vs. "easy" test;DDIF=Dummy for "difficult" test vs. "easy" test;MCON=Dummod * Cons interaction;DCON=Dumdif * Cons interaction;MCAR=Dummod * Care interaction;DCAR=Dumdif * Care. interaction;MMAT=Dummod * Mathanx interaction;DMAT=Dumdif * Mathanx interaction;MDANX=Dummod * Testanx interaction;DFANX=Dumdif * Testanx interaction;KWTST=Knowledge test

* - $p<.05$

Difficulty-based item characteristics and knowledge test scores

The first step in all of these analyses involved the entering of the two dummy variables corresponding to the three levels of difficulty-based item characteristics. The regression of percentage correct for the knowledge tests onto these two dummy variables yielded an R^2 value of .065. In repeated measures regression, however, only the within-subjects variance is relevant for within-subjects predictors. The R^2 value associated with the regression of the within-subjects variance in knowledge test scores onto the two within-subjects dummy variables was .20 ($F_{2,328}$=41.5, p<.01), suggesting that difficulty-based item characteristics had a substantial effect on knowledge test performance and, thus, providing support for Hypothesis 1.

Conscientiousness and knowledge test scores

The results with respect to the relationship between conscientiousness and knowledge test scores can also be found in Table 6.

Table 6

Regression of percentage correct on knowledge test onto conscientiousness and item characteristics

	Variable	W/in Ss	Btwn. Ss		
Step	Entered	S^2 explained	S^2 explained	df	F
1	Item Characteristics	.20		2, 328	41.5 ($p<.01$)
2	Conscientiousness		.00	1, 163	<1
3	Interaction	.015		2, 326	3.12 ($p<.05$)

As can be seen, conscientiousness had little effect on knowledge test scores. The conscientiousness by item characteristic interaction, however, was significant ($F_{2,326}$ =3.12, p<.05). This interaction is plotted in Figure 11. The plot was created by creating a regression equation for percentage correct with six components: one constant and five products. The five products represented each of the five terms (i.e., one term representing conscientiousness, two dummy variables representing difficulty, and two interactions between conscientiousness and the dummy variables) used in the regressions times their respective regression weights. When item difficulty was equal to 1 (ie., the " easy" test), the four terms involving dummy variables reduced to zero, leaving only the constant and the term containing the conscientiousness score. When item difficulty was equal to 2 (i.e., the " moderate" test), the terms involving the dummy variable corresponding to the "difficult" test reduced to zero. When item difficulty was equal to 3 (i.e., the "difficult" test), the terms involving the dummy variable corresponding to the "moderate" test reduced to zero.

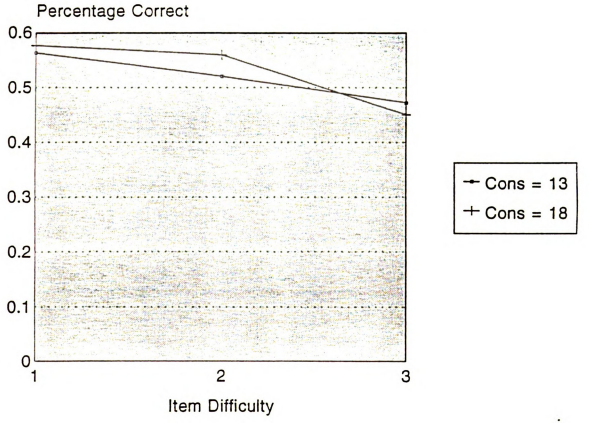

Figure 11. Plot of the effect on test scores of the conscientiousness by item characteristics interaction

Figure 11 shows that the interaction can be seen most clearly in the difference in the relationship between conscientiousness and percentage correct between the "moderate" and "difficult" tests. The plot shows that the interaction is not of the form expected. The knowledge test scores of respondents high in conscientiousness were more adversely affected by difficult item characteristics than were the scores of respondents low in Conscientiousness. The relationship between item difficulty and percentage correct was negative for both levels of conscientiousness. These analyses fail to provide support for analyses 2 and 6.

The results with respect to the relationship between carelessness and knowledge test scores can be found in Table 7.

Table 7

Regression of percentage correct on knowledge test onto carelessness and item characteristics

	Variable	W/in Ss	Btwn. Ss		
Step	Entered	S^2 explained	S^2 explained	df	F
1	Item Characteristics	.20		2, 328	41.5 ($p<.01$)
2	Carelessness		.21	1, 163	43.81 ($p<.01$)
3	Interaction	.002		2, 326	<1

Carelessness was significantly related to knowledge test scores, thus providing support for Hypothesis 3. The R^2 associated with the regression of between-subjects variance onto carelessness scores was .21 ($F_{1,163}$=43.81, p<.01). Those respondents who were higher in carelessness had lower knowledge test scores than did those respondents who were lower in carelessness. This effect, however, was not moderated by item characteristics, thus failing to provide support for Hypothesis 7. The R^2 value associated with this interaction was .0023.

Math anxiety and knowledge test scores

The results with respect to the relationship between math anxiety and knowledge test scores can be found in Table 8.

Table 8

Regression of percentage correct on knowledge test onto math anxiety and item characteristics

Step	Variable Entered	W/in Ss S^2 explained	Btwn. Ss S^2 explained	df	F
1	Item Characteristics	.20		2, 328	41.5 ($p<.01$)
2	Math Anxiety		.06	1, 163	9.87 ($p<.01$)
3	Interaction	.001		2, 326	<1

Math anxiety was significantly related to knowledge test scores. The R^2 associated with the regression of between-subjects variance onto math anxiety scores was .06 ($F_{1,163}$=9.87, p<.01). Those respondents who were higher in math anxiety had lower knowledge test scores than did those respondents who were lower in math anxiety as hypothesized (Hypothesis 4). This effect was not moderated by item characteristics, thus, Hypothesis 8 was not supported. The R^2 value associated with this interaction was .0023.

Test Anxiety and knowledge test scores

The results with respect to the relationship between test anxiety and knowledge test scores can be found in Table 9.

Table 9

Regression of percentage correct on knowledge test onto test anxiety and item characteristics

Variable		W/in Ss	Btwn. Ss		
Step	Entered	S^2 explained	S^2 explained	df	F
1	Item Characteristics	.20		2, 328	41.5 ($p<.01$)
2	Test Anxiety		.01	1, 163	2.35 ($p>.05$)
3	Interaction	.001		2, 326	<1

Neither test anxiety nor its interaction with item characteristics significantly predicted knowledge test scores. Test anxiety explained only 1.4% of the between subjects variance, and its interaction with item characteristics explained less than 1% of the within-subjects variance. Thus, there was no support for Hypotheses 5 and 9.

Difficulty-based item characteristics and l_z

As with the analyses of knowledge test scores, the first step in all of the analyses of l_z involved the entering of the two dummy variables corresponding to the three levels of difficulty-based item characterstics. The R^2 value associated with the regression of the within-subjects variance in l_z onto the two within-subjects dummy variables was .0005 ($F_{2,328}<1$), suggesting that difficulty-based item characteristics had no effect on l_z values.

Conscientiousness and l_z

The results with respect to the relationship between conscientiousness and l_z can be found in Table 10.

Table 10

Regression of l_i onto conscientiousness and item characteristics

Step	Variable Entered	W/in Ss S^2 explained	Btwn. Ss S^2 explained	df	F
1	Item Characteristics	.00		2, 328	<1
2	Conscientiousness		.002	1, 163	<1
3	Interaction	.007		2, 326	1.15 ($p>.05$)

Conscientiousness had a nonsignificant effect on l_z values ($F_{1,163}<1$), accounting for less than 1% of the between subjects variance in l_z. The effect of the interaction between item characteristics and conscientiousness on l_z was also nonsignificant (i.e., Hypothesis 10; $R^2=.007$; $F_{2,326}=1.15$; $p>.05$).

Carelessness and l_z

The results with respect to the relationship between conscientiousness and l_z can be found in Table 11.

Table 11

Regression of l, onto carelessness and item characteristics

	Variable	W/in Ss	Btwn. Ss		
Step	Entered	S^2 explained	S^2 explained	df	F
1	Item Characteristics	.00		2, 328	<1
2	Carelessness		.044	1, 163	7.50 ($p<.01$)
3	Interaction	.022		2, 326	3.61 ($p<.05$)

Carelessness had a significant effect on l_z values ($F_{1,163}$=7.50; p<.01), accounting for 4.4% of the between subjects variance. Those respondents who reflected more carelessness had lower l_z values (i.e., provided response patterns with greater inappropriateness) than did those respondents who reflected less carelessness. The effect for the carelessness by item characteristics interaction was also significant ($F_{2,326}$= 3.61;p<.05). This effect accounted for 2.2% of the within-subjects variance after removal of the relevant main effects. Figure 12 contains a plot of this interaction.

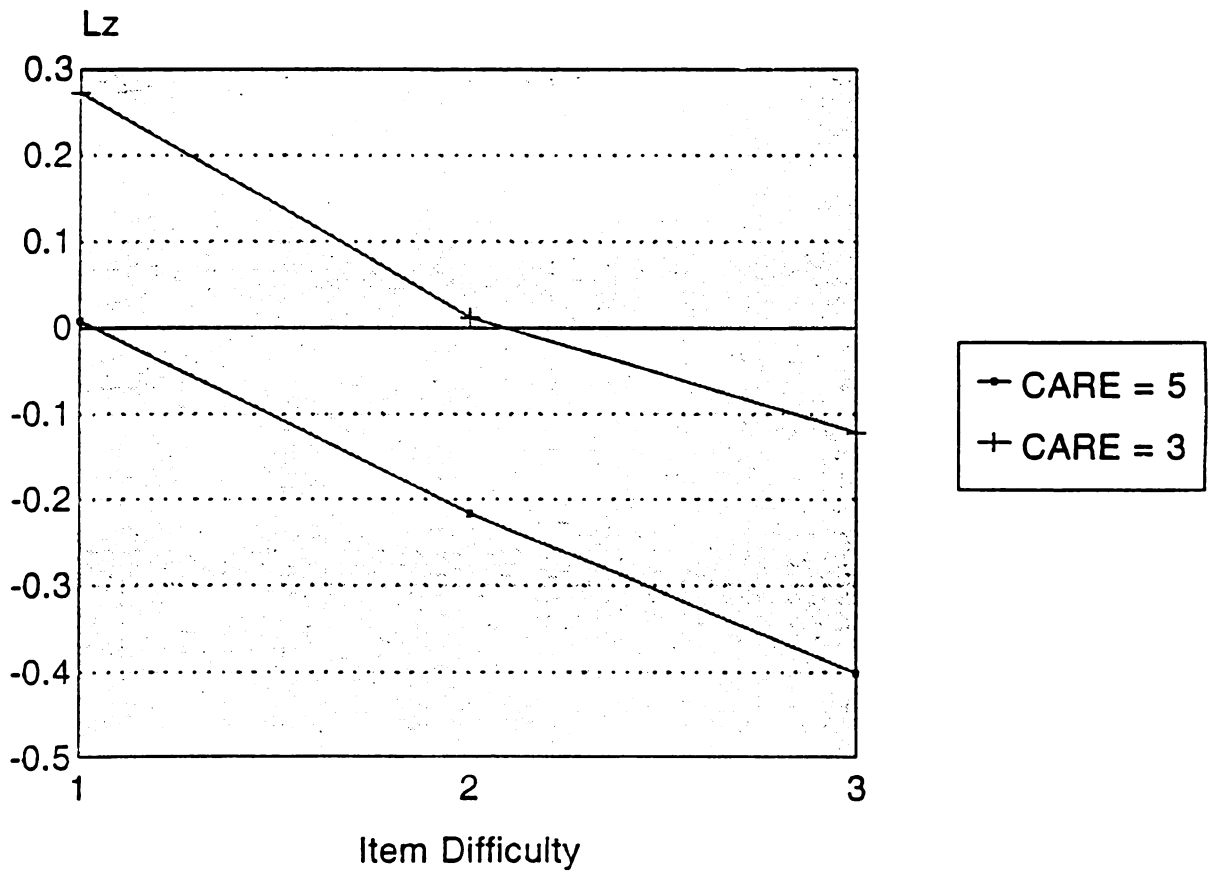

Figure 12. Plot of the effect on l_z of the interaction between carelessness and item characteristics

This interaction was also not of the form hypothesized (see Hypothesis 11). Those respondents who were higher in carelessness (i.e., respondents that had lower scores on the carelessness measure) displayed less inappropriateness as a function of item difficulty than did respondents lower in carelessness.

Math Anxiety and l_z

The results with respect to the relationship between conscientiousness and l_z can be found in Table 12.

Table 12

Regression of l, onto math anxiety and item characteristics

Step	Variable Entered	W/in Ss S^2 explained	Btwn. Ss S^2 explained	df	F
1	Item Characteristics	.00		2, 328	<1
2	Math anxiety		.008	1, 163	1.32 ($p>.05$)
3	Interaction	.008		2, 326	1.28 ($p>.05$)

Math anxiety did not have a significant effect on l_z values, accounting for less than 1% of the between-subjects variance ($F_{1,163}$=1.32; p>.05). The effect for the math anxiety by item characteristic interaction was also nonsignificant ($F_{2,326}$=1.28;p>.05). Thus, Hypothesis 12 was not supported.

Test anxiety

The results with respect to the relationship between conscientiousness and l_z can be found in Table 13.

Table 13

Regression of l, onto test anxiety and item characteristics

	Variable	W/in Ss	Btwn. Ss		
Step	Entered	S^2 explained	S^2 explained	df	F
1	Item Characteristics	.00		2, 328	<1
2	Test anxiety		.002	1, 163	<1
3	Interaction	.002		2, 326	<1

Test anxiety also failed to produce a significant effect on l_z values, accounting for less than 1% of the between-subjects variance ($F_{1,163}$=1.32; p>.05). The effect for the test anxiety by item characteristic interaction (Hypothesis 13) was also nonsignificant ($F_{2,326}$=1.28;p>.05).

Table 14 summarizes the findings of the present study with respect to the Hypotheses presented on p.66. As can be seen, support was found for Hypotheses 1, 3, and 4, which dealt with the effects of item characteristics, carelessness, and math anxiety on knowledge test scores. None of the other Hypotheses were supported.

Table 14

Summary of hypotheses and support

	Knowledge Test		l_z
	Main Effects	Interactions	Interactions
H1-Item Characteristics	X[a]		
H2-Conscientiousness	O		
H3-Carelessness	X		
H4-Math Anxiety	X		
H5-Test Anxiety	O		
H6-Conscientiousness		O	
H7-Carelessness		O	
H8-Math Anxiety		O	
H9-Test Anxiety		O	
H10-Conscientiousness			O
H11-Carelessness			O
H12-Math Anxiety			O
H13-Test Anxiety			O

[a] - X indicates support for the hypothesis while O indicates lack of support

Discussion

The purpose of the present study was to investigate the determinants of test inappropriateness. Specifically, I investigated the effects of difficulty-based item characteristics, math anxiety, test anxiety, conscientiousness, and carelessness on knowledge test scores and the l_z index of test inappropriateness. The following sections discuss the results of this study as they relate to the hypotheses that were presented on page 67.

Hypothesis 1: Difficulty-based item characteristics and test scores

Difficulty-based item characteristics had a profound effect on knowledge test scores. Specifically, difficulty-based item characteristics accounted for 20% of the within-subjects variance in knowledge test scores. As hypothesized, respondents chose the correct response option less often for items with many difficulty-based item characteristics than they did for items with fewer difficulty-based item characteristics. These results suggest that the responses to test items are a function not only of the standing of the respondent on the construct of interest, but also of the format of the item. This finding is consistent with previous research (e.g., Hughes & Trimble, 1965; Dudycha & Carpenter, 1973).

Hypothesis 2: Conscientiousness and test scores

Conscientiousness was found to have little effect on test scores, accounting for less than 1% of the between-subjects variance. Although contrary to hypothesis 2 of the present study, this is not entirely inconsistent with past research on the criterion-related validity of the Big Five personality dimensions. Meta-analyses of the relationship between conscientiousness and work-related outcomes have generally yielded uncorrected validities of less than .15 (Barrick & Mount, 1992; Tett et al., 1992).

Hypothesis 3: Carelessness and test scores

Carelessness was found to have a considerable effect on knowledge test scores, accounting for 21% of the between-subjects variance. As hypothesized, those respondents who exhibited a large degree of carelessness received lower test scores than did those who exhibited little or no carelessness. In fact, the mean knowledge scores ("easy" items, "moderate" items, and "difficult" items) for those respondents who answered all six carelessness items correctly were .62, .60, and .53 respectively whereas the scores for those respondents who missed at least one of the carelessness items were .50, .47, .39. This suggests that responses to test items were due not only to the standing of the respondent on the construct of interest, but also to the carelessness of the respondent at the time of test administration. In other words, there was evidence of Type P inappropriateness due to an effect for carelessness.

One possible alternative explanation for this finding is that the carelessness items used in the present study were in fact statistical knowledge items and, therefore, should be related to knowledge scores. While this is a possibility, it should be noted

that each of the items used in the carelessness scale were items which were answered correctly by all students who had responded to the item in previous tests in which those items had been used. Participants in the present study were still taking an Introductory statistics course at the time of testing, and the material contained in the carelessness items was material that had been covered in the class. So, while the content of the carelessness items was knowledge-related, the only viable cause of variance on the items (given that they are administered to people familiar with the rubric of statistics) was carelessness.

Hypothesis 4: Math anxiety and test scores

Math anxiety was found to have a significant effect on knowledge test scores, accounting for 6% of the between-subjects variance. As hypothesized, respondents higher in math anxiety had lower test scores than did respondents lower in math anxiety. This suggests that there was also evidence of Type P inappropriateness due to an effect for math anxiety.

Hypothesis 5: Test anxiety and test scores

Test anxiety was found to have little effect on test scores, accounting for less than 1% of the between-subjects variance. One explanation for this result is that the experimental testing situation lacked those aspects of real testing situations which lead to test anxiety. This possibility is discussed in more detail below.

Hypothesis 6: Conscientiousness by item characteristic interaction and test scores

Hypothesis 6 was not supported by the data. The conscientiousness by item characteristic interaction did contribute significantly to the prediction of test scores, but the form of the interaction was not as expected. It was hypothesized that

respondents low in conscientiousness would be more adversely affected by difficulty-based item characteristics than would respondents high in conscientiousness. Instead, the opposite was found. As can be seen in Figure 11, however, the extent of the interaction is slight and may have been due only to chance.

Hypothesis 7: Carelessness by item characteristic interaction and test scores

Hypothesis 7 was not supported by the data. The carelessness by item characteristic interaction did not contribute significantly to the prediction of test scores. In other words, there was no evidence of Type I inappropriateness resulting from an interaction between carelessness and item characteristics.

Hypothesis 8: Math anxiety by item characteristic interaction and test scores

Hypothesis 8 was not supported by the data. The math anxiety by item characteristic interaction did not contribute significantly to the prediction of test scores. In other words, there was no evidence of Type I inappropriateness resulting from an interaction between math anxiety and item characteristics.

Hypothesis 9: Test anxiety by item characteristic interaction and test scores

Hypothesis 9 was not supported by the data. The test anxiety by item characteristic interaction did not contribute significantly to the prediction of test scores. In other words, there was no evidence of Type I inappropriateness resulting from an interaction between math anxiety and item characteristics. As with the main effects, one explanation for this result is that the experimental testing situation lacked those aspects of real testing situations which lead to test anxiety.

Hypothesis 10: Conscientiousness by item characteristic interaction and l_z

Hypothesis 10 was not supported by the data. The conscientiousness by item characteristic interaction did not contribute significantly to the prediction of l_z. Although this is contrary to the hypothesis of the present study, it is not surprising given the lack of effect for this interaction on test scores. The lack of effect for the interaction on test scores suggests that there was little in the way of Type I inappropriateness (at least as caused by the conscientiousness by item characteristic interaction), therefore, any variance in l_z was due to other factors or chance.

Hypothesis 11: Carelessness by item characteristic interaction and l_z

Hypothesis 11 was not supported by the data. Although the carelessness by item characteristic interaction did contribute significantly to the prediction of l_z, the interaction was not of the form expected. It was hypothesized that respondents higher in carelessness would have higher l_z values only for the tests composed of items with difficulty-based item characteristics. Instead, respondents higher in carelessness displayed less of an item characteristic-l_z effect. So, while l_z was predicted by this interaction, it was not predicted in a way that was consistent with Hypothesis 11.

Hypothesis 12: Math anxiety by item characteristic interaction and l_z

Hypothesis 12 was not supported by the data. The math anxiety by item characteristic interaction did not contribute significantly to the prediction of l_z. As with conscientiousness, however, the lack of effect for this interaction on test scores suggests that there was little in the way of Type I inappropriateness (at least as caused by the math anxiety by item characteristic interaction), therefore, any variance in l_z was due to other factors or chance.

Hypothesis 13: Test anxiety by item characteristic interaction and l_z

Hypothesis 13 was not supported by the data. The test anxiety by item characteristic interaction did not contribute significantly to the prediction of test scores. In other words, there was no evidence of Type I inappropriateness resulting from an interaction between test anxiety and item characteristics. As with the main effects, one explanation for this result is that the experimental testing situation lacked those aspects of real testing situations which lead to test anxiety.

Implications and conclusions

The results of the present study have several implications for the interpretation of ability and knowledge test scores. First, item characteristics, in particular difficulty-based item characteristics, can have a profound effect on test scores. As was suggested in the Introduction, these item characteristics force the respondent to use abilities other than those related to the ostensible content of the items to arrive at the correct answer. Insofar as this is the case, the items which possess these characteristics are measuring constructs other than those that they are intended to measure.

The second implication for test scores has to do with carelessness. The present study found that carelessness had a considerable relationship with test scores. Specifically, the results suggest that those respondents who perform poorly on carelessness items answer far fewer items correctly than do those respondents who perform well on carelessness items. This is a particularly important issue for testing situations such as those encountered in concurrent, criterion-related validity studies in which respondents are existing employees who have nothing to gain from performing

well on the selection tests. If there is a significant number of respondents who are careless in their responding, then estimates of validity may be adversely affected. Also, if incumbent test scores are to be used in some fashion to set cutoffs for job applicants, carelessness may lead to cutoffs that are too low. In short, the results of the present study suggest that, in those situations where there are few or no formal rewards for test performance, test results can be interpreted in light of the effects of carelessness.

A third implication with respect to test scores is that math anxiety may have a considerable impact on math test scores, specifically, those respondents higher in math anxiety can be expected to answer fewer math knowledge questions correctly than respondents lower in math anxiety. The exact nature of this relationship, however, is not entirely clear. As with general test anxiety, a person may be high in math anxiety simply because of a lack of knowledge or ability (Naveh-Benjamin et al., 1987), in which case it would appear that the lack of knowledge is causing both the math anxiety and the lack of math performance. If, on the other hand, a respondent is high in math anxiety but does not lack the requisite knowledge or ability for a given task, then it is more likely that performance is determined by both anxiety and ability. More research is needed which isolates these factors so that their independent contributions to test performance can be identified.

To summarize, the results of the present study suggest that difficulty-based item characteristics such as response option complexity and negative wording and personality characteristics such as carelessness and math anxiety can have a substantial effect on math test scores. Future research should investigate the specific testing

situations in which these effects hold. Also, the present study focused only on statistical knowledge tests, which is an example of a test of maximum performance. Future research should endeavor to discover how these factors affect responses to tests of typical performance.

The present study also has implications for l_z. Interactions between item characteristics and each of the four personality variables were predicted, but none were found to be significant and in the hypothesized direction. One explanation for these findings is described in the Limitations section of this paper, namely, that a lack of external motivators in the testing situation led to uninterpretable results with respect to test scores. If this was the case, then l_z might also be uninterpretable.

Another possible explanation is that l_z simply does not reflect interactions between item characteristics and respondent characteristics. Instead, l_z might simply reflect nonsystematic response tendencies that, by definition, cannot be predicted. Before we accept this explanation, however, the relationship between l_z and interactions between item and respondent characteristics must be studied in testing situations which contain the external motivators that were missing in the present study. Only then can firm conclusions be drawn with respect to the hypotheses suggested in this paper.

Limitations

As was mentioned earlier, the most plausible explanation for the lack of hypothesized effects in the present study was that the testing situation lacked the external motivators present in most real-world, evaluation-oriented testing situations. Although efforts were made to encourage diligence on the part of the respondents,

grades, job opportunities, and other outcomes which can drive performance were not in any way contingent upon performance on the knowledge test. This lack of external motivators may have led to a failure to elicit reactions to the testing situation that would have been present if knowledge test performance were somehow tied to rewards. For example, most of the previous research on test anxiety has examined test anxiety in the context of an actual testing situation (i.e., research was not the only function served by the administration of the test). It may be that respondents who have an involuntary anxiety reaction to these true testing situations have little or no such reaction to a testing situation which lacks important consequences.

This lack of external motivators may also have led to a uniformly low level of effort on the knowledge test itself. Effects for variables such as conscientiousness might have been washed out by such a general lack of effort.

There are two final points to be made with respect to the issue of the role of external motivators in testing. First, while the testing situation used in the present study was unlike many real-world testing situations, it shared many of the characteristics of the typical concurrent validation study. In concurrent validation, as in the present study, participants respond to test items knowing that rewards are not contingent upon test performance. In fact, many such participants view the collection of concurrent data as a waste of their time. More research must be conducted which compares the effects of anxiety, conscientiousness, carelessness, and their interactions with item characteristics on item responses in situations which contain typical external motivators versus situations which do not contain such motivators. Such research

could shed more light on the nature of these predictors as well as on the nature of inappropriateness.

The second point to be made is related to the issue of the presence or absence of external motivators. It may be that a test such as the statistical knowledge test used in the present study simply isn't a valid measure of the construct of interest in a testing situation such as that used in the present study. It was suggested earlier that test appropriateness is an issue only for a test for which evidence of validity has been provided. If a certain level of effort is a prerequisite for the validity of a test, and if testing situations such as that used in the present study do not foster that level of effort, then perhaps appropriateness is not an issue in such situations. This is a question that the comparative research suggested above could address.

A final limitation of the present study is that it tested only part of the model presented on page 62. The relationships involving acquiescence, need for approval, field articulation, response sets, test wiseness, and omissiveness were not examined. The relationships involving acquiescence and need for approval are particularly promising for test of typical performance such as personality tests. As was discussed in the Introduction of this paper, acquiescence and need for approval may interact with item characteristics to affect test scores, and this interaction may be detectable with l_z. Likewise, field articulation and test wiseness show promise for tests measuring maximum performance. These factors may also interact with item characteristics to affect test scores, with the interaction being detected by l_z. As with the relationships examined in the present study, these relationships should be investigated in a testing situation which provides the external motivators typical of real-world testing situations.

LIST OF REFERENCES

Allport, G.W. (1928). A test for ascendance-submission. Journal of Abnormal Psychology, 23, 118-136.

Anastasi, A. (1988). Psychological Testing(6th ed.) New York: Macmillan.

Anderson, K.J. (1990). Arousal and the inverted-U hypothesis: A critique of Neiss's "Reconceptualizing arousal". Psychological Bulletin, 107, 96-100.

Benjamin, M., McKeachie, W.J., Lin, Y., & Holinger, D.P. (1981). Test anxiety: deficits in information processing. Journal of Educational Psychology, 73, 816-824.

Berg, I.A., & Collier, J.S. (1953). Personality and group differences in extreme response sets. Educational and Psychological Measurement, 13, 164-169.

Birenbaum, M. (1985). Comparing the effectiveness of several IRT-based appropriateness measures in detecting unusual response patterns. Educational and Psychological Measurement, 45, 523-535.

Block, J. (1965). The Challenge of Response Sets: Unconfounding Meaning, Acquiescence, and Social Desirability in the MMPI. New York: Irvington.

Bock, R.D., Dicker, C., & VanPelt, J. (1969). Methodological implications of content-acquiescence correlation in the MMPI. Psychological Bulletin, 71, 127-139.

Boring, E.G. (1950). A History of Experimental Psychology (rev. ed.). New York: Appleton, Century, Crofts.

Broverman, D.M., Klaiber, E.L., Kobayashi, Y., & Vogel, W. (1968). Roles of activation and inhibition in sex differences in cognitive ability. Psychological Review, 75, 23-50.

Cady, V.M. (1923). The estimation of juvenile incorrigibility. Journal of Delinquency Monograph, No.2.

Campeau, P.L. (1968). Test anxiety and feedback in programmed instruction. Journal of Educational Psychology, 59, 159-163.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.

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

Costa, P.T., & McCrae, R.R. (1988). From catalog to classification: Murray's needs and the five-factor model. Journal of Personality and Social Psychology, 55, 258-265.

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: agreeing response set as a personality variable. Journal of Abnormal Social Psychology, 60, 151-174.

Cronbach, L.J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.

Cronbach, L.J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.

Crowne, D.P., & Marlowe, D. (1964). The Approval Motive: Studies in Evaluative Dependence. New York: Wiley.Jackson, D.N., & Messick, S. (1958). content and style in personality assessment. Psycholigcal Bulletin, 55, 243-252.

Damarin, F. (1970). A latent structure model for answering personal questions. Psychological Bulletin, 73, 23-40.

Dolly, J.P., & Williams, K.S. (1986). Using test taking strategies to maximize multiple choice test scores. Educational and Psychological Measurement, 46, 619-625.

Donlon, T.F., & Fischer, F.E. (1968). An index of an individual's agreement with group-determined item difficulties. Educational and Psychological Measurement, 28, 105-113.

Drasgow, F., Levine, M.V., & Williams, E.A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. British Journal of Mathematical and Statistics Psychology, 38, 67-86.

Drasgow, F., Levine, M.V., & McLaughlin, M.E. (1991). Appropriateness measurement for some multidimensional test batteries. Applied Psychological Measurement, 15, 171-191.

Drasgow, F., Levine, M.V., Williams, B., McLaughlin, M.E., & Candell, G.L. (1989). Modeling incorrect responses to multiple-choice items with multilinear formula score theory. Applied Psychological Measurement, 13, 285-299.

Dreger, R.G., & Aiken, L.R. (1957). The identification of number anxiety in a college population. Journal of Educational Psychology, 48, 344-351.

DuBois, P.S. (1966). A test-dominated society: China 1115 B.C. - 1905 A.D. In A. Anastasi (Ed.), Testing Problems in Perspective (pp. 29-36). Washington, D.C.: American Council on Education.

Dudycha, A.L., & Carpenter, J.B. (1973). Effects of item format on item discrimination and difficulty. Journal of Applied Psychology, 58, 116-121.

Edwards, A.L. (1957). The Social Desirability Variable in Personality Assesement and Research. New York: Dryden.

Endler, N.S., & Hunt, J.M. (1966). Sources of behavioral variance as measured by the S-R inventory of anxiousness. Psychological Bulletin, 65, 336-346.

Fagley, N.S. (1987). Positional response bias in multiple-choice tests of learning: its relationship to test wiseness and guessing strategies. Journal of Educational Psychology, 79, 95-98.

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.

Forehand, G.A. (1962). Relationships among response sets and cognitive behaviors. Educational and Psychological Measurement, 22, 287-302.

Gaier, E.L., Lee, M.C., & McQuitty, L.L. (1953). Response patterns in a test of logical inference. Educational and Psychological Measurement, 13, 550-567.

Gehman, W.S. (1957). A study of ability to fake scores on the Strong Vocational Interest Blank for men. Educational and Psychological Measurement, 17, 65-70.

Gibb, B.G. (1964). Test Wiseness as Secondary Cue Response. Unpublished doctoral dissertation, Stanford University.

Goodenough, D.R. (1976). The role of individual differences in field dependence as a factor in learning and memory. Psychological Bulletin, 83, 675-694.

Green, R.F. (1951). Does a selection situation induce testees to bias their answers on interest and temperament tests? Educational and Psychological Measurement, 11, 503-515.

Harnisch, D.L., & Linn, R.L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 18, 133-146.

Harper, F.B.W. (1974). The comparative validity of the Mandler-Sarason Test Anxiety questionnaire and the Achievement Anxiety Test. Educational and Psychological Measurement, 34, 961-966.

Hartshorne, H., & May, M.A. (1928). Studies in Deceit. New York: Macmillan.

Henry, E.M., & Rotter, J.B. (1956). Situational influences on Rorschach responses. Journal of Consulting Psychology, 20, 457-462.

Herrmann, D.J. (1982). Know thy memory: The use of questionnaires to assess and study memory. Psychological Bulletin, 92, 434-452.

Hollenbeck, J. R., Ilgen, D.W., & Sego, D. (in press). Repeated measures regression and mediational tests: Enhancing the power of leadership research. Leadership Quarterly.

Hothersall, D. (1990). History of Psychology (2nd Ed.) New York: McGraw-Hill.

Hough, L.M., Eaton, N.K. Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75, 581-595.

Hughes, H.H., & Trimble, W.E. (1965). The use of complex alternatives in multiple choice items. Educational and Psychological Measurement, 25, 117-126.

Humm, D.G., & Humm, K.A. (1944). Validity of the Humm-Wadsworth temperament scale: with consideration of the effects of subject's response-bias. Journal of Psychology, 18, 55-64.

Humm, D.G., Storment, R.C., & Iorns, M.E. (1939). Combination scores for the Humm-Wadsworth temperament scale. Journal of Psychology, 7, 227-253.

Jackson, D.N. & Messick, S. (1958). Content and style in personality assessment. Psychological Bulletin, 55, 245-252.

Jackson, D.N. & Messick, S. (1965). Acquiescence: the nonvanishing variance component. American Psychologist, 20, 498-502.

Kingston, A.J., George, C.E., & Ewens, W.P. (1956). Determining the relationship between individual interest profiles and occupational forms. Journal of Educational Psychology, 47, 310-316.

Klein, G.S., Barr, H.C., & Wolitzky, D.C. (1967). Personality. Annual Review of Psychology, 18, 467-560.

Lawrence, P.J. (1957). Some characteristics of incorrect responses to intelligence test items. Australian Journal of Psychology, 9, 1-11.

Levine, M., & Rubin, D. (1979). Measuring the appropriateness of multiple choice test scores. Journal of Educational Statistics, 4, 269-290.

Lorge, I., & Diamond, L.K. (1954). The prediction of absolute item difficulty by ranking and estimating techniques. Educational and Psychological Measurement, 14, 2-10.

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

Messick, S. (1966). The Psychology of Acquiescence: A Interpretation of Research Evidence. Educational Testing Sevice: Princeton, NJ, April.

Metfessel, N.S., & Sax, G. (1957). Response set patterns in published Intructor's Manuals in education and psychology. California Journal of Educational Research, 8, 195-197.

Metfessel, N.S., & Sax, G. (1958). Systematic biases in the keying of correct responses on certain standardized tests. Educational and Psycholigcal Measurement, 18, 787-790.

Millman, J., Bishop, C.H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Naveh-Benjamin, M., McKeachie, W.J., & Lin, Y.G. (1987). Two types of test anxious students: support for an information processing model. Journal of Educational Psychology, 79, 131-136.

Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: Further support for the information processing model. Journal of Educational Psychology, 83, 134-139.

Paulman, R.G., & Kennally, K.J. (1984). Test anxiety and ineffective test takers: Different names, same construct? Journal of Educational Psychology, 76, 279-288.

Peabody, D. (1966). Authoritarianism scales and response bias. Psychological Bulletin, 65, 11-23.

Peterson, R.A. (1961). A technique for the detection of blind checking in questionnaire research. Educational and Psychological Measurement, 21, 361-362.

Rapaport, G.M., & Berg, I.A. (1955). Response sets in a multiple-choice test. Educational and Psychological Measurement, 15, 58-62.

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Kobenhavn: Danmarks Paedagogiske Institut. (Reprinted by University of Chicago Press, 1980).

Reise, S.P. (1990). A comparison of item and person-fit methods of assessing model-data fit in IRT. Applied Psychological Measurement, 14, 127-137.

Rorer, L.G. (1965). The great response style myth. Psychological Bulletin, 63, 129-156.

Rosenberg, N., Izard, C.E., & Hollander, E.P. (1955). Middle category response: reliability and relationship to personality and intelligence variables. Educational and Psychological Measurement, 15, 281-290.

Rosenzweig, S. (1934). A suggestion for making verbal personality test more valid. Psychological Review, 41, 400-401.

Ruch, F.L. (1942). A technique for detecting attempts to fake performance on a self-inventory type of personality test. In Q. McNemar and M.A. Merrill, Studies in Personality. (pp. 229-234), New York: McGraw-Hill.

Rudner, L.M. (1983). Individual assessment accuracy. Journal of Educational Measurement, 20, 207-219.

Sarnacki, R.E., (1979). An examination of testwiseness in the cognitive test domain. Review of Educational Research, 2, 252-279.

Sato, T. (1975). The Construction and Interpretation of S-P Tables. Meiji Toosho.

Satterly, D.J., (1976). Cognitive styles, spatial ability, and school achievement. Journal of Educational Psychology, 68, 36-42.

Saupe, J.L. (1960). An empirical model for the corroboration of suspected cheating on multiple-choice tests. Educational and Psychological Measurement, 20, 475-489.

Schmitt, N., Gooding, R.Z., Noe, N.A., & Kirsch, M. (1984). Metaanalyses of validity studies between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schuman, H., & Kalton, G. (1985). Survey methods. In G. Lindzey and E. Aronson (Eds.) Handbook of Social Psychology. Hillsdale, NJ: Erlbaum.

Tatsuoka, K.K. & Tatsuoka, M.M. (1980). Detection of aberrant response patterns and their effect on dimensionality. Research Report 80-4-ONR Urbana, Il: University of Illinois, Computer based Education Research Laboratory.

Tatsuoka, K.K. & Tatsuoka, M.M. (1982). Detection of aberrant response patterns. Journal of Educational Statistics, 7, 215-231.

Tobias, S. (1979). Anxiety research in educational psychology. Journal of Educational Psychology, 71, 573-582.

vanderFlier, H. (1977). Environmental factors and deviant response patterns. In Y.H. Poortinga (Ed.) Basic Problems in Cross Cultural Psychology. Amsterdam: Swets & Seitlinger, B.V.

Wahsltrom, M., & Boersma, F.J. (1968). The influence of test-wiseness upon achievement. Educational and Psychological Measurement, 28, 413-420.

Waterhouse, I.K. & Child, I.L. (1953). Frustration and the quality of performance: An experimental study. Journal of Personality, 22, 298-311.

Watson, D, & Clark, L.A. (1984). Negative affectivity: the disposition to experience negative emotional states. Psychological Bulletin, 96, 465-490.

Whitcomb, M.A., & Travers, R.M.W. (1957). A study of within-test learning functions as a determinant of total score. Educational and Psychological Measurement, 17, 86-97.

Wiggins, J.S. (1962). Strategic, method, and stylistic variance in the MMPI. Psychological Bulletin, 59, 224-242.

Witkin, H.A. (1974). Cognitive Styles, Essence and Origins: Field Dependence and Field Independence. New York: International Universities Press.

Wright, B.D., & Panchapakesan, N.A. (1969). A procedure for sample free item analysis. Educational and Psychological Measurment, 29, 23-48.

Zimmerman, W.S. (1954). The influence of item complexity upon the factor composition of a spatial visualization test. Educational and Psychological Measurement, 14, 106-119.

LIST OF APPENDICES

APPENDIX A

Personality measures

Conscientiousness

Please use the following scale to answer questions 1 - 4:

A - Strongly disagree
B - Disagree
C - Neutral
D - Agree
E - Strongly Agree

Please mark your answers on the bubble sheet that was provided.

I try to perform all of the tasks assigned to me conscientiously.

When I make a commitment, I can always be counted on to follow through.

I am a productive person who always gets the job done.

I strive for excellence in everything I do.

Test Anxiety

Please use the following scale to answer questions 5 - 15:

A- The statement does not describe my present condition
B - The condition is barely noticeable
C - The condition is moderate
D - The condition is strong
E - The condition is very strong; the statement describes my present condition well.

I feel my heart beating fast.

I feel regretful.

I am so tense that my stomach is upset.

I am concerned about others in the class seeing the results of this test.

I have an uneasy, upset feeling.

I feel that others will be disappointed in me.

I am nervous.

I feel I may not do as well on this test as I could.

I feel panicky.

I do not feel very confident about my performance on this test.

Math Anxiety

When my stats professor (or any math professor) asks questions to find out how much we know about a particular mathematical concept or approach, I worry that I will do poorly in the class.

When my stats professor is showing the class how to do a particular problem, I worry that the other students in the class migh understand the problem better than I do.

When I am in stats (or any math class), I usually feel relaxed and at ease.

When I am taking a math test, I usually feel relaxed and at ease.

Taking math tests scares me.

I dread having to do math.

The thought of taking a more advanced stats class (e.g., Psych 302, 304, or required stats courses in graduate school) scares me.

In general, I worry about how well I am doing in school.

If I miss a given day of stats class, I worry that I will be behind the other students when I come back.

In general, I worry about how well I am doing in math.

Compared to other subjects, I worry a great deal about how well I am doing in math.

Carelessness Items

In a one-way ANOVA, if the degrees of freedom between groups is equal to the number of groups minus 1, and there are 4 groups, what would the degrees of freedom between groups be?

a. 6
b. 7
c. 3
d. 8

Which of the following does ANOVA stand for?

a. ANalysis Of VAriance
b. Standard Deviation
c. Correlation
d. Repeated Measures Designs

The two types of errors that one can make in the sort of hypothesis testing exemplified by the above scenario are

a. Type I and Type II
b. Type III and Type IV
c. Type V and Type VI
d. Type VII and Type VIII

Which of the following are within the range of possible correlations?

a. -.50
b. .40
c. .50
d. all of the above

(This question does not apply to the scenario that the previous questions applied to or any other particular scenario) In general, which of the following are within the range of possible z-scores?

a. 1.0
b. 2.0
c. -1.0
d. all of the above

Another word for the average is

a. mean
b. variance
c. sample
d. population

APPENDIX B

Statistical knowledge items

For questions 1 - 75, use the following scenario.

Suppose I am interested in examining the effect of the number of times one watches "Jeopardy" and the number of times one watches "Wheel of Fortune" on the number of Trivial Pursuit questions that one can answer. In an attempt to do this, I assign 30 people to one of two levels of Jeopardy-watching and one of three levels of Wheel of Fortune-watching. The data look like this.

		JEOPARDY WATCHING	
		NONE	10 TIMES
	NONE	58 75 68 37 27 --- 265	45 51 75 72 69 --- 312
WHEEL OF FORTUNE WATCHING	10 TIMES	61 41 55 50 47 --- 254	55 58 42 68 32 --- 255
	20 TIMES	75 65 50 35 30 --- 255	31 31 46 32 40 --- 180

(Diff) 1.) If each subject had received every level of one of the variables in the "Trivial Pursuit" experiment, I could not use

a. a three-way ANOVA
b. a Chi-squared test
c. a one-way ANOVA
d. all of the above

(Mod) 2.) If I had simply wanted to compare the number of people who watch Jeopardy to the number of people who watch Wheel of Fortune, I would have to use which of the following?

a. a three-way ANOVA
b. a Chi-squared test
c. a one-way ANOVA
d. any of the above

For questions 3-8, use the following scenario.

Suppose I am interested in the effects of the number of Jagermeister shots that one does on the number of questions that one can answer about analysis of variance. In an attempt to assess these effects, I assign 15 people to one of three groups so that there are 5 people in each group. GROUP 1 gets no Jager, GROUP 2 gets 3 shots of Jager, and GROUP 3 gets 8 shots of Jager. The data look like this:

GROUP 1	GROUP 2	GROUP 3
3	2	6
7	7	8
3	8	8
1	2	7
0	4	8

$X^2 = 482$

(Diff) 3.) A statement that does not represent the null hypothesis for this test is

a. $_1 = \ _2 = \ _3$
b. $_1 = \ _2$
c. $_1 = \ _2 = \ _3$
d. both a and b

(Diff) 4.) The degrees of freedom Between, Within, and Total for this test are not

a. 3, 11, 14
b. 2, 13, 15
c. 2, 12, 14
d. both a and b

(Mod) 5.) What are the Sums of Squares Between, Within, and Total?

a. 58.8, 61.2, 120
b. 53.73, -63.2, 116.93
c. 53.73, 63.2, 116.93
d. none of the above

(Mod) 6.) If the Sums of Squares Between and Within are 53.73 and 63.2 respectively (they aren't necessarily), what are the Mean Squares Between and Within?

a. 17.91, 5.74
b. 26.86, 5.27
c. 3.58, 4.21
d. none of the above

(Mod) 7.) Which of the following is true of the F-ratio?

a. $F_{2,14} = 5.10$
b. $F_{2,12} = 5.10$
c. $F_{2,12} = .25$
d. none of the above

(Mod) 8.) With which of the following levels of significance would you reject the null?

a. .05
b. .01
c. .005
d. both a and c

(Easy) 9.) A Mean Square is a form of which of the following?

a. standard error
b. variance
c. standard deviation
d. covariance

(Easy) 10.) Which of the following represents a factorial design?

a. Each level of one variable is paired with one level of every other variable
b. Each variable is paired with every other variable
c. One variable is paired with every level of one other variable
d. Each level of every variable is paired with each level of every other variable

For questions 11-12, use the following scenario

Suppose I am interested in assessing the effects of job satisfaction (IV1) and salary (IV2) on job performance (DV). So, I collect data on these three variables for 100 people and correlate the variables. The correlation between satisfaction and performance is .23, the correlation between salary and performance is .45, and the correlation between salary and satisfaction is .42.

(Mod) 11.) Which of the following is the type of regression analysis that would I use to assess the effects of job satisfaction and salary on job performance?

a. multiple regression
b. repeated measures regression
c. simple regression
d. any of them depending on the nature of the variables

(Mod) 12.) Which of the following is the regression equation?

a. $Y = .05X_1 + .43X_2 + a$
b. $Y = .07X_1 + .61X_2$
c. $Z_y = .05Z_1 + .43Z_2$
d. none of the above

(Easy) 13.) In general, the F-ratio will be large if which of the following is true?

a. If the variance within the groups is larger than the variance between the groups
b. If the variance within the groups is equal to the variance between the groups
c. If the variance within the groups is less than the variance between the groups
d. If there is no variance between the groups

(Easy) 14.) Which of the following best represents the null hypothesis for a one-way ANOVA with three groups?

a. $_1 - _2 = 0 \quad _1 - _3 = 0 \quad _2 - _3 = 0$
b. $_1 - _2 = 0$
c. $_1 = _2$
d. $_1 = _3$

(Easy) 15.) Which of the following is true of ANOVA?

a. It is relatively insensitive to violations of the normality assumption but not the homogeneity assumption
b. It is relatively insensitive to violations of the homogeneity assumption but not the normality assumption
c. It is relatively insensitive to violations of both of its assumptions
d. It is greatly affected by violations of either of its assumptions

(Easy) 16.) Which of the following has to be true in order for the Central Limit Theorem to be applicable?

a. the samples come from the same population
b. the samples come from different populations
c. the sample means are equal
d. the population variances are different

(Easy) 17.) If I wanted to examine the effects in an ANOVA further, which of the following should I use?

a. a second ANOVA
b. an acid test
c. urinalysis
d. multiple comparison procedures

For questions 18 - 19, use the following scenario.

Suppose I am interested in knowing the difference in ACT scores between men and women. In an attempt to investigate this, I draw a sample of 16 men and a sample of 16 women. The mean of the ACT scores for the men is 18.6 and the mean for the women is 20.4. The variances of the individual scores are 6.1 and 8.9 respectively. Suppose further that the standard error for the differences between means is .97.

(Mod) 18.) Calculate a t-score for the difference between these two means. (When using the t-score formula, be sure to put the means in the order that they were presented)

a. 1.85
b. -1.85
c. -1.8
d. none of the above

(Diff) 19.) I would fail to reject the null with

a. a one-tailed test that examines the lower 5% of the distribution
b. a two-tailed test with a significance level of .05.
c. a two-tailed test with a significance level of .10.
d. both a and b

Use the following scenario to answer questions 20 - 21

Suppose I am interested in knowing the difference in ACT scores between men and women. In an attempt to investigate this, I draw five samples of men with 12 in each sample and 5 samples of women with 12 in each sample. For each pair of samples, I record the mean ACT score for men, the mean ACT score for women, and the difference between the means so that my data look like this

ACT (MEN)	ACT (WOMEN)	DIFFERENCE
$X_{11} = 18$	$X_{21} = 24$	$X_{11} - X_{21} = -6$
$X_{12} = 22$	$X_{22} = 22$	$X_{12} - X_{22} = 0$
$X_{13} = 17$	$X_{23} = 23$	$X_{13} - X_{23} = -6$
$X_{14} = 20$	$X_{24} = 26$	$X_{14} - X_{24} = -6$
$X_{15} = 16$	$X_{25} = 27$	$X_{15} - X_{25} = -11$
$X_{x1} = 18.6$	$X_{x2} = 24.4$	
$s^2_{x2} = 5.8$	$s^2_{x2} = 4.3$	

(Mod) 20.) If we draw another sample of 15 men and 15 women, and their mean ACT scores are 17 and 22 respectively, and the standard error of the differences between means is 3.18, what is the t-score for the difference between these two means?

a. 1.73
b. -1.57
c. .49
d. none of the above

(Mod) 21.) What would the effect on the power of my t-test be if I doubled my sample size?

a. My power would increase
b. My power would decrease
c. My power would be unaffected
d. Either a or b depending on the situation

For questions 22 - 24, use the following scenario.

Suppose I am interested in knowing the difference in armpit hair between men and women. In an attempt to investigate this, I draw a sample of 21 men and a sample of 21 women. The mean for pithair for men is 4.6 and the mean for women is 6.6. The variances of the individual scores within these samples are 7.1 and 7.7 respectively.

(Diff) 22.) The standard error of differences is not
a. .70
b. .84
c. 14.8
d. This applies to both a and c

(Mod) 23.) Calculate the t-score for the difference between these two means.
a. -2.38
b. -2.0
c. 2.0
d. none of the above

(Mod) 24.) With which of the following would I fail to reject the null?
a. a one-tailed test that examines the upper 5%
b. a one-tailed test that examines the lower 1%
c. a two-tailed test with a significance level of .05
d. both a and b

For questions 25 - 26, use the following scenario

Suppose I want to estimate the average number of hours per day that Dan Quayle spends playing with his Legos. In order to do this, I take a sample of 12 days and record the number of hours that he spends playing with his Legos on each of those days. They are as follows: 4, 3.5, 6, 7, 3.5, 4.5, 6.5, 8, 1.5, 2.5, 6, 5. The standard deviation of this set of scores is 1.93.

(Mod) 25.) What is the 95% confidence interval around the mean? (Hint: You must use the value for a two-tailed test)
a. .58 < < 9.08
b. 3.6 < < 6.06
c. 3.09 < < 6.57
d. none of the above

(Mod) 26.) What is the 99% confidence interval around the mean? (Hint: You must use the value for a two-tailed test)

a. .58 < < 9.08
b. 3.6 < < 6.06
c. 3.09 < < 6.57
d. none of the above

(Easy) 27.) What is a confidence interval?

a. It is a range of scores within which we expect the population mean to fall.
b. It is the population mean
c. It is a range of scores within which we expect the sample mean to fall.
d. It is our two best estimates of the population mean.

(Diff) 28.) Suppose I was interested in knowing the relationship between goal difficulty and goal commitment (a fairly common topic in my field). In an attempt to investigate this relationship, I draw a sample of 15 people, give them a goal for a task, and get the correlations between goal difficulty and commitment. The correlation from these 15 people is -.48. The t-score for this correlation would not be

a. -1.73
b. 1.73
c -1.97
d. b and c

For questions 29 - 33, use the following scenario

Suppose I know that the mean of the population of grade point averages at MSU is 2.35. Suppose further that I get a sample of GPA's from an unknown source, and the scores are: 3.3, 3.6, 2.2, 2.4, 3.1, 1.6, 1.4, 2.9, 3.9, 2.5, 3.4, 3.3. The mean of these scores is 2.8 and the standard deviation is .79.

(Diff) 29.) The t-score for this mean is not

a. .45
b. .57
c. -1.97
d. all of the above

(Mod) 30.) What would the degrees of freedom for this t-test be?

a. 12
b. 11
c. 132
d. either a or b depending on the situation

(Diff) 31.) We would fail to reject the null with
a. a one-tailed test which examines the lower 5% of the distribution
b. a two-tailed test with a level of significance of .05
c. a two-tailed test with a level of significance of .01
d. all of the above

(Mod) 32.) Which of the following would happen if the sample size were doubled?
a. The t-score that I calculate would be larger and the score that I use from the Table would be smaller
b. The t-score that I calculate would be smaller and the score that I use from the Table would be larger.
c. The t-score would be larger and the score from the Table would be unaffected
d. Neither the calculated t nor the score from the Table would be affected

(Diff) 33.) If the sample variance were doubled, it is not the case that
a. the t-score that I calculate would be larger and the score that I use from the Table would be smaller
b. the t-score that I calculate would be smaller and the score that I use from the Table would be larger.
c. the t-score would be smaller and the score from the Table would be unaffected
d. Both a and c apply

(diff) 34.) Suppose I was interested in knowing the relationship between goal difficulty and goal commitment (a fairly common topic in my field). In an attempt to investigate this relationship, I draw a sample of 15 people, give them a goal for a task, and get the correlation between goal difficulty and commitment. The correlation from these 15 people is -.48. The degrees of freedom for the t-test for correlations is not
a.) 13
b.) N-1
c.) N-2
d.) This applies to none of the above

(Easy) 35.) Which of the following is true of an unstandardized regression weight (i.e., b in a regression equation with a single predictor)?
a. It is the amount of change in the dependent variable associated with a unit change in the independent variable
b. It is the amount of change in the independent variable associated with a unit change in the dependent variable
c. It is the correlation between the independent and dependent variables
d. It is the covariance between the independent and dependent variables

(Diff) 36.) Suppose I know the number of times each Michigan resident has been swindled by Gov. Engler (So, I have access to this population of scores). I then take many different samples of 15 people each and calculate the mean for each sample. If the mean of the means were 7.8 and the standard deviation of individual scores were 2.2, the population mean and the standard error of the mean would not be

a. = 2.79, = .57

b. = 7.8, = 2.2

c. either a or b

d. all of the above

(Diff) 37.) I am interested in predicting graduate school GPA from GRE scores. In an attempt to do this, I take a sample of 10 graduate students, record their GRE scores and their GPA's. The data look like this

subject #	GRE (X)	Grad School GPA (Y)
1	730	2.9
2	1120	3.1
3	1310	3.9
4	810	3.2
5	960	3.0
6	1250	3.5
7	1180	3.3
8	1410	3.7
9	840	3.4
10	660	3.1

If the correlation between these two variables is .75, it could not be said that GRE scores account for _____

a. 68% of the variance in GPA
b. 95% of the variance in GPA
c. 57% of the variance in GPA
d. both a and b

(Diff) 38.) I am interested in examining the relationship between GRE scores and GPA in graduate school. In an attempt to do this, I take a sample of 10 graduate students, record their GRE scores and their GPA's. The data look like this: (by the way, GRE scores can range from 400 - 1600)

subject #	GRE (X)	Grad School GPA (Y)
1	1230	3.4
2	1120	3.1
3	1310	3.4
4	1210	3.9
5	1360	3.8
6	1250	3.5
7	1180	3.3
8	1410	3.6
9	1380	3.4
10	1100	3.6

If the covariance between these two variables is 7.44, this (without any knowledge of the variances) would not tell you

a. anything
b. that there is a strong, positive relationship
c. that there is a weak, positive relationship
d. This applied to none of the above

For questions 39 - 40, use the following scenario.

Suppose I am interested in assessing the relationship between schizophrenia (measured with a 10 point scale) and job performance (measured on a 10 point scale). To do this, I collect data on both of these variables for 30 people.

(Mod) 39.) If the scatterplot for the data looked like the following, what kind of relationship would it be?

a. imperfect negative
b. imperfect positve
c. perfect positive
d. none of the above

(Diff) 40.) If the scatterplot looked like the following, it would not be a(n)

a. imperfect negative relationship
b. imperfect positve relationship
c. positive relationship
d. Both b and c apply

(Easy) 41.) Which of the following correlation coefficients represents the strongest relationship?

a. .68
b. .22
c. -.46
d. -.82

(Easy) 42.) Which of the following correlation coefficients represents the weakest relationship?

a. .68
b. .22
c. -.46
d. -.82

(Easy) 43.) Which of the following covariances represents the strongest relationship?

a. 4.28
b. 17.71
c. -24.94
d. can't tell

(Easy) 44.) What is the probability of rolling a 1 then a 2 then a 3 then a 4 in four rolls of a die?

a. .005
b. .51
c. .00077
d. .67

(Easy) 45.) What is the probability of rolling a 1 or a 2 or a 3 or a 4 in a single roll of a die?

a. .005
b. .51
c. .00077
d. .67

(Easy) 46.) What is the probability of rolling a 1 then a 2 or a three then a four in two rolls of a die?

a. .11
b. .055
c. .67
d. .00077

(Diff) 47.) Suppose I know that the mean of the population of grade point averages at MSU is 2.35. Suppose further that I get a sample of GPA's from an unknown source, and the scores are: 3.3, 3.6, 2.2, 2.4, 3.1, 1.6, 1.4, 2.9, 3.9, 2.5, 3.4, 3.3. The mean of these scores is 2.8 and the standard deviation of these individual scores is .79. If I knew that the null hypothesis were actually false (This would never actually happen), and I used a one-tailed test that examined the lower 5% of the distribution, I would not commit

a. a Type 1 error
b. a Type 2 error
c. an error
d. either a or c

For questions 48 - 50, use the following scenario

Suppose I know that the mean number of "beauty surgeries" that Phyllis Schlafly has in a day is 3.1 (a population mean), and the standard deviation of these individual scores (sigma) is 1.25. I then receive data on the number of surgeries that an unknown nazi fraulein has each day for ten days. The mean for these ten days is 3.7.

(Diff) 48.) The z-score for this mean is not

a. 4.8
b. 1.54
c. .6
d. both a and c

(Mod) 49.) What is the null hypothesis for this test?

a. The number of beauty surgeries that Phyllis Schlafly has is from the sample for the unknown nazi
b. The sample is from the population of Phyllis Schlafly surgeries
c. The sample has a mean that is larger than the population mean
d. none of the above

(Diff) 50.) If I knew that the null hypothesis were actually true (This would never actually happen), and I used a one-tailed test that examined the upper 5% of the distribution, I would not commit

a. a Type 1 error
b. a Type 2 error
c. an error
d. all of the above

For questions 51 - 52, use the following scenario.

Suppose I measured the height (in inches) of MSU football players and found that they were normally distributed with a mean of 75 and a standard deviation of five.

(Diff) 51.) The percentage of MSU football players that can be expected to be between 70 and 80 inches tall is not

a. 68%
b. 95%
c. roughly two-thirds
d. This applies to none of the above

(Mod) 52.) Approximately what percentage of MSU football players can be expected to be between 65 and 85 inches tall?

a. 68%
b. 95%
c. one-third
d. either a or b

(Mod) 53.) What would the Z-score be for an MSU football player who was 83 inches tall?

a. 1.6
b. 16
c. .16
d. none of the above

(Diff) 54.) Referring to the Z-score that you just calculated, the percentage of the heights that you would expect not to fall above this score is

a. 5%
b. 95%
c. either a or b
d. none of the above

(Easy) 55.) Which of the following is the term used to describe the extent to which the scores in a set of scores congregate in the tails of the distribution?

a. skew
b. positive skew
c. potato stew
d. kurtosis

(Easy) 56.) Which of the following is the term used to describe the extent to which a distribution is asymmetrical?

a. skew
b. positive skew
c. potato stew
d. kurtosis

(Easy) 57.) Which of the following is a property of the normal distribution?

a. It is skewed
b. It is unimodal
c. It is leptokurtic
d. It is platykurtic

(Easy) 58.) The variance is basically which of these?

a. the average squared deviation from the mean
b. the sum of the squared deviations from the mean
c. the sum of the absolute deviations from the mean
d. the average absolute deviation from the mean

(Easy) 59.) Which of the following is an advantage of the mean as a measure of central tendency?

a. It is greatly affected by extreme scores
b. It can be manipulated algebraically
c. It is not greatly affected by extreme scores
d. It is difficult to calculate

(Easy) 60.) Which of the following is a disadvantage of the mean as a measure of central tendency?

a. It is affected by extreme scores
b. It can be manipulated algebraically
c. It is not greatly affected by extreme scores
d. It is difficult to calculate

(Easy) 61.) Consider the following set of numbers:

4, 15, 7, 5, 1, 5, 6, 8

Using these numbers, calculate the percentile for a score of 8.

a. .125
b. 8.75
c. .7
d. .875

For questions 62 - 63, use the following scenario.

Suppose I wish to assess people's height and political affiliation. To do this, I take 100 people and record their height in inches. I also assign them a 1 if they are democrat, 2 if republican, and 3 if they are other.

(Diff) 62.) Height in inches is not a
- a. nominal scale
- b. ordinal scale
- c. interval scale
- d. all of the above

(Mod) 63.) What type of scale would political affiliation be?
- a. nominal
- b. ordinal
- c. interval
- d. none of the above

(Easy) 64.) Which of the following is clearly an interval scale?
- a. Height in inches
- b. Height in centimeters
- c. Temperature (in Celsius)
- d. Gender

(Easy) 65.) Which of the following is clearly a ratio scale?
- a. Weight in Kilograms
- b. Class standing
- c. Grade Point Average
- d. Temperature (in Celsius)

For questions 66 - 69, use the following scenario.

For some strange and terrible reason, I am interested in knowing the average number of white collar crimes committed per day by Sen. Bob Dole over the past 2000 days. In an attempt to estimate this value, I randomly choose twenty days from these 2000, count the number of white collar crimes he committed on each of those 20 days, and get the average of those twenty numbers.

(Diff) 66.) My sample size is not
- a. twenty
- b. the average for the twenty days
- c. fifty Million
- d. either b or c

(Diff) 67.) The parameter that I am trying to estimate is not
a. the average number of white collar crimes per day committed by Dole over the past 2000 days.
b. the average number of white collar crimes committed per day by Dole over the 20 days that I measured.
c. the total number of white collar crimes committed by Dole divided by 2000
d. This applies to none of the above

(Mod) 68.) What is the statistic that I have used?
a. The average number of white collar crimes per day committed by Dole over the past 2000 days.
b. The average number of white collar crimes committed per day by Dole over the 20 days that I measured.
c. 2000
d. all of the above

(Mod) 69.) What is the population of interest?
a. The white collar crimes committed per day by Dole over the past 2000 days
b. The white collar crimes committed per day by Dole over the past 20 days
c. The average number of white collar crimes per day committed by Dole over the past 2000 days.
d. Either a or c.

Use the following scenario for questions 70 - 74.

For some strange and terrible reason, I am interested in knowing the average height of the fifty million people who voted for Ross Perot (the second coming of Thurston Howell). In an attempt to estimate this value, I find twenty people who voted for Perot, measure their heights, and calculate their average.

(Mod) 70.) What is my population of interest?
a. The cast of Gilligan's Island
b. The twenty people whose heights I measured
c. All voters
d. none of the above

(Mod) 71) What is my sample size?
a. Twenty
b. Not enough information provided
c. Fifty Million
d. none of the above

(Diff) 72.) The group that is not the sample that I am using is
a. the fifty million who voted for Perot
b. the twenty people whose heights I measured
c. both a and b
d. none of the above

(Diff) 73.) The numerical value that is not the parameter that I am trying to estimate is
a. The average height of the fifty million people who plan to vote for Ross Perot
b. The average height of the twenty people that I measured.
c. Fifty Million
d. both b and c

(Mod) 74.) What is the statistic that I have used?
a. The average height of the fifty million people who plan to vote for Ross Perot
b. The average height of the twenty people that I measured.
c. Fifty Million
d. either a or b

(Easy) 75.) Which of the following is the term for numerical values used to make generalizations about a large set of data from a subset of that set?
a. descriptive statistics
b. dependent statistics
c. inferential statistics
d. parametric statistics

APPENDIX C

Descriptives for statistical knowledge items

Variable	Mean	Std Dev	Minimum	Maximum	N	Label
ITEMA1	.73	.44	0	1	165	
ITEMA2	.58	.49	0	1	165	
ITEMA3	.92	.27	0	1	165	
ITEMA4	.64	.48	0	1	165	
ITEMA5	.38	.49	0	1	165	
ITEMA6	.65	.48	0	1	165	
ITEMA7	.75	.44	0	1	165	
ITEMA8	.33	.47	0	1	165	
ITEMA9	.67	.47	0	1	165	
ITEMA10	.25	.44	0	1	165	
ITEMA11	.28	.45	0	1	165	
ITEMA12	.75	.43	0	1	165	
ITEMA13	.71	.46	0	1	165	
ITEMA14	.23	.42	0	1	165	
ITEMA16	.63	.48	0	1	165	
ITEMA17	.67	.47	0	1	165	
ITEMA18	.32	.47	0	1	165	
ITEMA19	.54	.50	0	1	165	
ITEMA20	.33	.47	0	1	165	
ITEMA21	.44	.50	0	1	165	
ITEMA22	.84	.37	0	1	165	
ITEMA23	.46	.50	0	1	165	
ITEMA24	.54	.50	0	1	165	
ITEMA25	.67	.47	0	1	165	
ITEMA26	.27	.45	0	1	165	
ITEMA28	.78	.41	0	1	165	
ITEMA29	.27	.44	0	1	165	
ITEMA30	.15	.35	0	1	165	
ITEMA31	.27	.44	0	1	165	
ITEMA32	.61	.49	0	1	165	
ITEMA33	.75	.43	0	1	165	
ITEMA34	.45	.50	0	1	165	
ITEMA35	.68	.47	0	1	165	
ITEMA36	.68	.47	0	1	165	
ITEMA37	.60	.49	0	1	165	
ITEMA38	.58	.49	0	1	165	
ITEMA39	.32	.47	0	1	165	
ITEMA40	.17	.38	0	1	165	
ITEMA41	.36	.48	0	1	165	
ITEMA42	.28	.45	0	1	165	
ITEMA44	.19	.40	0	1	165	
ITEMA45	.68	.47	0	1	165	
ITEMA46	.88	.33	0	1	165	
ITEMA47	.43	.50	0	1	165	
ITEMA48	.44	.50	0	1	165	
ITEMA49	.35	.48	0	1	165	
ITEMA50	.65	.48	0	1	165	
ITEMA51	.50	.50	0	1	165	
ITEMA52	.55	.50	0	1	165	
ITEMA53	.54	.50	0	1	165	
ITEMA54	.63	.48	0	1	165	
ITEMA55	.70	.46	0	1	165	
ITEMA56	.49	.50	0	1	165	
ITEMA57	.39	.49	0	1	165	
ITEMA59	.43	.50	0	1	165	
ITEMA60	.64	.48	0	1	165	
ITEMA61	.47	.50	0	1	165	

ITEMA62	.39	.49	0	1	164
ITEMA64	.70	.46	0	1	165
ITEMA65	.47	.50	0	1	165
ITEMA66	.73	.44	0	1	165
ITEMA67	.68	.47	0	1	165
ITEMA69	.52	.50	0	1	165
ITEMA70	.12	.32	0	1	165
ITEMA71	.52	.50	0	1	165
ITEMA72	.48	.50	0	1	165
ITEMA73	.63	.48	0	1	165
ITEMA74	.62	.49	0	1	165
ITEMA75	.46	.50	0	1	165
ITEMA76	.43	.50	0	1	165
ITEMA77	.53	.50	0	1	165
ITEMA78	.48	.50	0	1	165
ITEMA79	.35	.48	0	1	165
ITEMA81	.30	.46	0	1	165
ITEMA82	.12	.32	0	1	165

- - Correlation Coefficients - -

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6	ITEM7	ITEM8	ITEM9	ITEM10	ITEM11
ITEMA1	1.0000	.1000	.1797*	.1994*	.1566*	.0807	.1825*	.0409	.1051	.0063	.1610*
ITEMA2	.1000	1.0000	.1625*	.0232	.0996	-.0474	.0969	.0938	.1681*	.0441	.0887
ITEMA3	.1797*	.1625*	1.0000	.2466**	-.0982	.1187	.1906*	.0122	.3234**	.0160	.1317
ITEMA4	.1994*	.0232	.2466**	1.0000	.0922	.0602	.1946*	.1513	.1172	-.0789	.1328
ITEMA5	.1566*	.0996	-.0982	.0922	1.0000	-.1469	.1374	.0989	-.0722	-.0512	.0200
ITEMA6	.0807	-.0474	.1187	.0602	-.1469	1.0000	.1314	.1264	.4168**	.1904*	.1674*
ITEMA7	.1825*	.0969	.1906*	.1946*	.1374	.1314	1.0000	.0518	.1855*	-.0099	.1771*
ITEMA8	.0409	.0938	.0122	.1513	.0989	.1264	.0518	1.0000	.1286	.1262	.0560
ITEMA9	.1051	.1681*	.3234**	.1172	-.0722	.4168**	.1855*	.1286	1.0000	.0814	.1456
ITEMA10	.0063	.0441	.0160	-.0789	-.0512	.1904*	-.0099	.1262	.0814	1.0000	-.0220
ITEMA11	.1610*	.0887	.1317	.1328	.0200	.1674*	.1771*	.0560	.1456	-.0220	1.0000
ITEMA12	.2558**	.3086**	.1442	.0610	.0697	.0542	.2757**	.1918*	.1668*	.0784	.2324**
ITEMA13	.1871*	.0251	.1099	.1261	-.1368	-.0444	.0546	-.0367	.0652	.0067	.1006
ITEMA14	.1020	.0844	.1065	.0843	-.0083	.1249	-.0439	.0786	.0134	.1100	.2899**
ITEMA16	.2479**	.0889	.0090	.2041**	.2053**	.1301	.1289	.1596*	.1883*	.0728	.0842
ITEMA17	.1344	.1157	.0357	.0903	.1411	.0909	.1855*	.0736	.2568**	-.0965	.0592
ITEMA18	.1213	.1096	-.0397	.0343	.0291	.1176	.0742	.1841*	.0925	-.0146	.0644
ITEMA19	.1027	.1286	.1811*	.2114**	-.1115	.0958	.1020	.0744	.2106**	.0376	.0593
ITEMA20	.0701	-.1157	-.0357	.0439	-.0078	.0721	.0814	.0916	-.0090	.0372	.0560
ITEMA21	.1713*	.0523	-.0148	.1062	.0743	.0224	.2617**	.0635	.0668	.0469	.1070
ITEMA22	.1037	.1232	.1139	.0743	.0049	-.0113	.1176	.0990	.0755	-.0800	.0193
ITEMA23	.1173	-.0547	.0446	.0414	-.0140	.0321	-.0183	.0551	-.0292	.1020	.1034
ITEMA24	.0477	.2765**	.1359	.0850	.0140	.1213	.1299	.1004	.1847*	.0655	.1949*
ITEMA25	.1551*	.0782	.0795	.0267	.0442	-.0811	.1181	.1096	.0274	.0000	.1242
ITEMA26	.0615	.0226	.0276	.0952	.1430	-.0702	.0454	.0659	.0211	.0795	-.0166
ITEMA28	.0796	.1471	.2268**	.0582	-.0446	.0791	.1292	.1183	.1006	.0055	.1321
ITEMA29	.0847	.0667	.0746	.1709*	-.1000	.1499	.0692	.0175	.1577*	.0566	.0835
ITEMA30	.0544	.1058	-.0070	.0260	.0348	.0467	.1227	.0420	-.0420	.0746	.0118
ITEMA31	.0227	.0667	.1255	-.0285	-.0151	.0058	-.0252	.0175	.0701	.0252	.0224
ITEMA32	-.0019	.1825*	.1365	.1998*	-.0501	.0756	-.0083	.0781	.1075	.0369	.1343
ITEMA33	-.0296	.1380	.0401	.0318	-.0172	.0247	.0181	.0723	.1967*	-.0181	.0447
ITEMA34	.0550	.0830	.2218**	-.0690	.1211	.0745	-.0533	.0896	.0920	.0813	.0296
ITEMA35	.1722*	.0746	.0397	.3164**	.1049	.1280	.1641*	.1202	.0458	.1338	.0803
ITEMA36	.2694**	.0861	.0921	.3008**	.0415	.1930*	.2325**	.1395	.1385	.1568*	.0726
ITEMA37	-.0168	.0602	-.0092	.0000	-.0562	.1093	.0057	-.0105	.1160	-.0909	.1214
ITEMA38	.2112**	.1282	-.0199	.1254	-.0018	-.0733	.0405	.0938	.0371	-.1251	-.1031
ITEMA39	.0256	.0726	.0531	-.0567	.0663	.0264	.0969	.0273	.1395	.0229	.0437
ITEMA40	.0170	.1214	.0123	.0732	.0160	.1247	.1159	-.0400	.1433	.0323	.1870*
ITEMA41	.1925*	.0172	.0774	.0382	.0217	.1697*	.1166	.0186	.1161	-.0296	.0720
ITEMA42	.2288**	.0995	.1347	.1142	.1480	.0349	.1530*	.2180**	.0682	.0011	.2065**
ITEMA44	-.1202	-.1435	-.0841	-.0116	-.1590*	-.1272	-.1356	-.1788*	-.2132**	.0301	-.0999

ITEMA45	.0548	.1273	-.0085	.0736	.1585*	.0189	.1344	.0372	.0734	.0742	.1382
ITEMA46	.0700	-.0513	-.1086	.1825*	.0964	.0817	.0388	.0612	-.1007	-.0814	.0238
ITEMA47	.1366	.0171	.0724	.1226	.0587	-.0379	.1425	.0721	-.0199	.0823	.0875
ITEMA48	-.0147	.1862*	.0340	-.0369	.0647	.0569	.0443	.0548	.0752	.0397	.1537*

	ITEMA1	ITEMA2	ITEMA3	ITEMA4	ITEMA5	ITEMA6	ITEMA7	ITEMA8	ITEMA9	ITEMA0	ITEMA11
ITEMA49	.0421	.1352	-.0203	.0238	.1627*	.0010	.0514	.0275	-.0817	-.1097	.0518
ITEMA50	.1301	-.1095	-.0740	.0240	.0732	.0257	.0652	-.0005	-.0266	.0223	.0897
ITEMA51	.0311	.1157	-.0207	.0550	.1204	.0171	.0870	-.0042	.0817	.0799	.0773
ITEMA52	-.0202	.0013	-.0828	.0023	.0958	-.0401	.1445	-.1242	.0203	.0234	.0987
ITEMA53	.1301	.2026**	.0908	-.0414	-.0613	-.0321	.1020	.0485	.0551	.0376	.1136
ITEMA54	.1623*	.2416**	.0556	-.1091	.0757	.2093**	.0136	.0258	.0545	.0728	.0282
ITEMA55	.1180	.1481	.0069	.0326	.1482	-.0259	.0770	.1424	.0293	-.0465	.1083
ITEMA56	.0713	.0952	.1072	.1375	.1393	.0760	.1564*	.1419	.0648	-.0172	.0383
ITEMA57	.0374	.0046	.0516	.0422	-.0109	.0901	.1010	.0721	.0072	-.1010	.0796
ITEMA59	.1089	.1412	.0724	.1226	.0587	-.0122	.1144	-.0583	.0199	.0542	.0875
ITEMA60	.0570	.0743	.0128	-.0214	.0402	.0072	.1946*	.1245	.0708	-.0210	.1328
ITEMA61	.0421	.2020**	.0932	.0505	.0518	.0664	.1283	.2278**	.0052	-.0167	.1770*
ITEMA62	.0330	.1247	.1422	.0627	.0981	.0852	.0971	.0352	.0285	.0175	.1683*
ITEMA64	.0280	.0944	.1053	.0602	-.0161	.0857	.0465	.1424	.1120	-.0161	.1674*
ITEMA65	-.0604	.0398	.0967	-.0413	-.0829	-.1035	-.0598	-.1430	.0898	.0319	.0881
ITEMA66	.0393	.1000	.1289	.0285	.0717	.0519	.1825*	.0701	.0175	-.0252	.0693
ITEMA67	.0629	.0596	-.0047	.0838	.2031**	.1382	.1427	.0561	.0829	-.0828	.1599*
ITEMA69	.0805	-.0501	.1700*	.1582*	.1174	.1457	.0527	.1514	.0813	.0309	.0818
ITEMA70	.0458	-.0406	.0350	.0359	-.1231	.0225	-.0507	-.0898	.0088	.0071	-.0126
ITEMA71	.1280	-.0358	.0314	.0733	.2018**	-.1182	.0177	-.0728	.0564	-.1291	-.1270
ITEMA72	.0640	-.0380	-.1214	.1535*	.1487	-.0348	.0936	.0728	.0470	-.0101	.0459
ITEMA73	.0208	.1398	.0090	.0736	.1276	-.0547	.1577*	-.0010	.0277	-.0136	-.0278
ITEMA74	.1830*	.0018	.0053	.0378	.0077	.0943	.0924	.2211**	.0456	-.0350	.1754*
ITEMA75	.1998*	.1425	.0897	.1425	.0362	.0065	.1771*	.1329	.0485	.1578*	.0763
ITEMA76	.0812	.0916	.2087**	-.0046	.0081	-.0637	.0301	.0721	.0583	.0542	.0056
ITEMA77	.1776*	.0197	.1323	.1263	-.0017	.1124	.1785*	.1346	.1502	.0725	-.0415
ITEMA78	.1390	.0747	.0551	.0940	-.0172	.0584	.1980*	-.0997	.0038	-.1423	.0535
ITEMA79	.0346	.1250	.0705	-.0072	.0416	-.0619	.0442	.0909	.0365	-.1319	.0315
ITEMA81	.0696	-.0024	-.1009	-.0498	.0602	-.0756	-.0385	.0179	.1303	.0083	.0312
ITEMA82	.0029	.0364	.0350	.0359	-.0839	-.0574	.0800	.0316	.0721	.0071	-.0549

	ITEM12	ITEM13	ITEM14	ITEM16	ITEM17	ITEM18	ITEM19	ITEM20	ITEM21	ITEM22	ITEM2
ITEMA1	.2558**	.1871*	.1020	.2479**	.1344	.1213	.1027	.0701	.1713*	.1037	.1173
ITEMA2	.3086**	.0251	.0844	.0889	.1157	.1096	.1286	-.1157	.0523	.1232	-.0547
ITEMA3	.1442	.1099	.1065	.0090	.0357	-.0397	.1811*	-.0357	-.0148	.1139	.0446
ITEMA4	.0610	.1261	.0843	.2041**	.0903	.0343	.2114**	.0439	.1062	.0743	.0414
ITEMA5	.0697	-.1368	-.0083	.2053**	.1411	.0291	-.1115	-.0078	.0743	.0049	-.0140
ITEMA6	.0542	-.0444	.1249	.1301	.0909	.1176	.0958	.0721	.0224	-.0113	.0321
ITEMA7	.2757**	.0546	-.0439	.1289	.1855*	.0742	.1020	.0814	.2617**	.1176	-.0183
ITEMA8	.1918*	-.0367	.0786	.1596*	.0736	.1841*	.0744	.0916	.0635	.0990	.0551
ITEMA9	.1668*	.0652	.0134	.1883*	.2568**	.0925	.2106**	-.0090	.0668	.0755	-.0292
ITEM10	.0784	.0067	.1100	.0728	-.0965	-.0146	.0376	.0372	.0469	-.0800	.1020
ITEM11	.2324**	.1006	.2699**	.0842	.0592	.0644	.0593	.0560	.1070	.0193	.1034
ITEM12	1.0000	.0949	.1147	.1697*	.1967*	.0351	.2565**	.1022	.2232**	.1627*	.1093
ITEM13	.0949	1.0000	.2236**	.0623	.1220	.0120	.0239	.0770	.0793	.1135	.0029
ITEM14	.1147	.2236**	1.0000	.1505	.0134	.1170	.1012	.0173	-.0169	.0863	.0721
ITEM16	.1697*	.0623	.1505	1.0000	.6699**	.0160	.1487	.1328	.1675*	.1024	.0024
ITEM17	.1967*	.1220	.0134	.6699**	1.0000	.1202	.2624**	.1562*	.1709*	.1454	.0485
ITEM18	.0351	.0120	.1170	.0160	.1202	1.0000	.1149	-.0649	.1014	.2341**	.0413
ITEM19	.2565**	.0239	.1012	.1487	.2624**	.1149	1.0000	.0744	.0776	.1829*	.0489
ITEM20	.1022	.0770	.0173	.1328	.1562*	-.0649	.0744	1.0000	.4541**	.0292	-.1522
ITEM21	.2232**	.0793	-.0169	.1675*	.1709*	.1014	.0776	.4541**	1.0000	.2240**	.0450
ITEM22	.1627*	.1135	.0863	.1024	.1454	.2341**	.1829*	.0292	.2240**	1.0000	-.0843
ITEM23	.1093	.0029	.0721	.0024	.0485	.0413	.0489	-.1522	.0450	-.0843	1.0000
ITEM24	.2002**	.0774	.1300	.0731	.1070	.2190**	.0974	-.0033	.1021	.0843	-.0730
ITEM25	.2182**	.1415	.1425	.0178	.1644*	.0734	.2235**	.1918*	.1815*	.2085**	.0860
ITEM26	.0687	.0626	.1499	.0461	.0211	.0159	-.0347	.1529*	.1472	.0134	.0074
ITEM28	.2735**	.1140	.0450	-.0398	.1319	.1748*	.1890*	.0870	.1097	.2027**	.0760

ITEM29	.2199**	-.0060	.1259	.1779*	.1577*	.0548	-.0752	.1051	.0774	.0074	.1027
ITEM30	.0781	.0372	-.0624	.0311	.0313	.1212	.0019	.0053	.1569*	-.0498	.0326
ITEM31	-.0338	-.0362	.0282	.0927	.0409	-.0333	.0898	.1051	.0497	.0445	-.0623
ITEM32	.0028	.0652	.0514	.0603	-.0516	.2013**	.1128	-.0545	.0233	.0513	.1367
ITEM33	.2211**	.0640	-.0186	-.0046	.0473	.0351	.0032	.0424	.0252	.0868	.1937*
ITEM34	.0179	-.0853	.1945*	.0688	.0920	.1801*	.0622	-.0920	.0558	-.0568	.1088
ITEM35	.2052**	.0166	.2530**	.1722*	.0734	.1119	.1715*	.2032**	.2389**	.1869*	-.0153
ITEM36	.2439**	.0825	.2471**	.0750	-.0283	.1314	.1583*	.1673*	.2286**	.1231	.0249
ITEM37	.1603*	-.0599	-.0823	-.0615	.0105	.0318	.2134**	.0105	.1447	.0067	.0099
ITEM38	.1949*	.0251	.0260	.1398	.1943*	.1622*	.1040	.0152	.1761*	-.0097	.0439
ITEM39	.1184	-.1687*	.0627	.0871	.0839	.0362	-.0013	-.0561	.0607	.0885	.0536
ITEM40	.1105	-.0659	.1745*	.1790*	.2121**	.1731*	.2234**	-.0056	.0580	.0690	.1005
ITEM41	.1656*	.0324	.0424	.2046**	.1431	.1096	.2328**	-.0622	.0320	.1591*	.2239**
ITEM42	.2075**	-.0688	.1651*	.0661	-.0177	.2273**	.0713	.0749	.1216	.0977	.0364
ITEM44	-.0017	.0779	.0229	-.1959*	-.1806*	-.1405	-.0388	.0499	-.1534*	-.1145	-.0227
ITEM45	.1451	-.0120	.0680	.2529**	.1841*	.1952*	.1715*	.0372	.1604*	-.0938	-.0674
ITEM46	.0872	-.0335	.1149	.0233	.0576	.0566	.0666	.0216	.0647	-.0137	.1569*
ITEM47	.1032	.0446	.1351	.1584*	.1627*	.0575	.1646*	.0460	.0745	.0866	.0564
ITEM48	.1734*	.0870	.1214	.1514	.2052**	.1973*	.1132	.0288	-.0702	.0642	-.0642

	ITEMA12	ITEMA13	ITEMA14	ITEMA16	ITEMA17	ITEM18	ITEM19	ITEM20	ITEM21	ITEM22	ITEM23
ITEM49	.0121	.0244	.1400	.0116	-.0815	.0644	-.0073	-.0807	-.0847	.0512	.0327
ITEM50	.0760	.0595	.0409	.0147	.0817	.1259	.1091	.0807	-.0177	.0861	.1201
ITEM51	.1297	.0573	.1406	.2432**	.2626**	.1386	.0542	.0216	.1169	-.0137	.0430
ITEM52	.0173	.0932	.0302	.0667	.1242	.0201	-.0999	.0316	.0809	-.0695	.0754
ITEM53	.2283**	.0506	.1300	.1487	.1847*	.1409	.1950*	.0226	.2002**	.0514	.0977
ITEM54	.0245	-.0483	.0014	.0377	.0010	.1235	-.0024	-.0812	.0410	-.1012	.1536*
ITEM55	.1174	.0218	.0405	.0518	.2251**	.1346	.1445	-.1120	-.0433	.0711	.1482
ITEM56	.1438	.0150	.0963	.1242	.0390	.1553*	.0562	.0644	-.0573	-.0244	.1627*
ITEM57	-.0244	.0521	.0598	.0265	-.0192	.1892*	-.1259	-.1394	.0409	.0213	.0264
ITEM59	.1315	.0985	.1351	.1838*	.1888*	-.0998	-.0073	-.0323	.0745	-.0457	.1301
ITEM60	.1484	.0151	.0245	.0214	.1709*	.1153	.1356	.1245	.0808	.1424	-.0345
ITEM61	.0600	-.0695	.1520	-.0638	.0311	.1110	.0845	-.0570	.0833	.0854	.0374
ITEM62	.1342	.0201	.0347	.0725	.1083	.1264	.0917	-.0817	.0479	.0518	.1840*
ITEM64	.3015**	.0510	.0405	.0793	.0838	.0778	.1977*	.0858	.1439	.0352	.0684
ITEM65	.0388	-.0617	.0587	-.0041	-.0381	-.0534	.1444	.0898	-.0988	.0251	-.0469
ITEM66	.1290	-.0543	.0369	.1060	.1928*	.1800*	.1301	.0117	.0332	-.0074	-.0202
ITEM67	.0929	-.0611	.1232	.2912**	.2775**	.0755	.0536	.0561	.0971	.1231	.0772
ITEM69	.1227	.0806	.1497	-.0052	.0813	.0098	.0879	.2548**	.0605	.1664*	-.0392
ITEM70	-.1001	-.1452	-.0620	-.0777	-.1530*	.0771	-.0856	.0316	.0654	.0056	-.1048
ITEM71	.0315	-.0607	-.0742	.1111	.1245	-.0858	-.0206	.0822	.0711	.0626	-.0280
ITEM72	-.0315	-.0194	-.0122	.1903*	.1598*	-.0700	.0450	.0211	.0267	-.0298	.0280
ITEM73	.0245	-.0483	-.0880	.2457**	.1883*	-.0109	.0227	.0525	.0156	.1024	.0024
ITEM74	.2489**	-.0010	.0975	.0539	.0989	.1049	.1617*	.0344	.0266	-.0387	.1646*
ITEM75	.1656*	.0029	.0432	.1284	.0744	.0934	.0977	.1070	.0941	.0472	.0242
ITEM76	.1315	.1524	.1351	.0317	.1627*	.0837	.0664	.0199	-.0489	.0866	.1301
ITEM77	.3055**	-.0642	.0212	.0638	.1243	.1492	.1836*	.0828	.1862*	.0788	.2064**
ITEM78	.0177	.0529	.0520	-.0702	-.0296	.0942	.0581	.0555	.2086**	.1616*	.0149
ITEM79	.0638	-.0117	-.0039	.0283	.1264	.1280	.0576	-.0178	-.0224	.0802	-.0321
ITEM81	.0740	-.0422	.0152	.0952	.0102	.1677*	-.0257	.0179	-.0483	-.0292	-.1595*
ITEM82	-.0122	.0220	-.1522	-.0384	-.0721	-.0042	-.0476	-.1302	-.0494	.0569	.0069

	ITEM24	ITEM25	ITEM26	ITEM28	ITEM29	ITEM30	ITEM31	ITEM32	ITEM33	ITEM34	ITEM35
ITEM1	.0477	.1551*	.0615	.0796	.0847	.0544	.0227	-.0019	-.0296	.0550	.1722*
ITEM2	.2765**	.0782	.0226	.1471	.0667	.1058	.0667	.1825*	.1380	.0830	.0746
ITEM3	.1359	.0795	.0276	.2268**	.0746	-.0070	.1255	.1365	.0401	.2218**	.0397
ITEM4	.0850	.0267	.0952	.0582	.1709*	.0260	-.0285	.1998*	.0318	-.0690	.3164**
ITEM5	.0140	.0442	.1430	-.0446	-.1000	.0348	-.0151	-.0501	-.0172	.1211	.1049
ITEM6	.1213	-.0811	-.0702	.0791	.1499	.0467	.0058	.0756	.0247	.0745	.1280
ITEM7	.1299	.1181	.0454	.1292	.0692	.1227	-.0252	-.0083	.0181	-.0533	.1641*
ITEM8	.1004	.1096	.0659	.1183	.0175	.0420	.0175	.0781	.0723	.0896	.1202
ITEM9	.1847*	.0274	.0211	.1006	.1577*	-.0420	.0701	.1075	.1967*	.0920	.0458
ITEM10	.0655	.0000	.0795	.0055	.0566	.0746	.0252	.0369	-.0181	.0813	.1338
ITEM11	.1949*	.1242	-.0166	.1321	.0835	.0118	.0224	.1343	.0447	.0296	.0803
ITEM12	.2002**	.2182**	.0687	.2735**	.2199**	.0781	-.0338	.0028	.2211**	.0179	.2052**

ITEM13	.0774	.1415	.0626	.1140	-.0060	.0372	-.0362	.0652	.0640	-.0853	.0166
ITEM14	.1300	.1425	.1499	.0450	.1259	-.0624	.0282	.0514	-.0186	.1945*	.2530**
ITEM16	.0731	.0178	.0461	-.0398	.1779*	.0311	.0927	.0603	-.0046	.0688	.1722*
ITEM17	.1070	.1644*	.0211	.1319	.1577*	.0313	.0409	-.0516	.0473	.0920	.0734
ITEM18	.2190**	.0734	.0159	.1748*	.0548	.1212	-.0333	.2013**	.0351	.1801*	.1119
ITEM19	.0974	.2235**	-.0347	.1890*	-.0752	.0019	.0898	.1128	.0032	.0622	.1715*
ITEM20	-.0033	.1918*	.1529*	.0870	.1051	.0053	.1051	-.0545	.0424	-.0920	.2032**
ITEM21	.1021	.1815*	.1472	.1097	.0774	.1569*	.0497	.0233	.0252	.0558	.2389**
ITEM22	.0843	.2085**	.0134	.2027**	.0074	-.0498	.0445	.0513	.0868	-.0568	.1869*
ITEM23	-.0730	.0860	.0074	.0760	.1027	.0326	-.0623	.1367	.1937*	.1088	-.0153
ITEM24	1.0000	.0430	.1564*	.2184**	.1448	.2088**	.0348	.2126**	.0877	.0866	.2236**
ITEM25	.0430	1.0000	.0289	.1868*	.0775	.0365	-.0678	-.0088	.0694	.1291	.0092
ITEM26	.1564*	.0289	1.0000	.0599	-.0308	-.0211	.0000	.0127	.1002	.1242	.0715
ITEM28	.2184**	.1868*	.0599	1.0000	.1195	.1347	-.0796	.1818*	.1716*	-.0188	.0137
ITEM29	.1448	.0775	-.0308	.1195	1.0000	-.0155	-.0847	.1144	.0930	.0826	.2094**
ITEM30	.2088**	.0365	-.0211	.1347	-.0155	1.0000	.0622	.0109	.0383	-.0314	.0629
ITEM31	.0348	-.0678	.0000	-.0796	-.0847	.0622	1.0000	.0300	-.0655	.1376	.0333
ITEM32	.2126**	-.0088	.0127	.1818*	.1144	.0109	.0300	1.0000	.2906**	.2271**	-.0149
ITEM33	.0877	.0694	.1002	.1716*	.0930	.0383	-.0655	.2906**	1.0000	.1306	-.0351
ITEM34	.0866	.1291	.1242	-.0188	.0826	-.0314	.1376	.2271**	.1306	1.0000	-.0237
ITEM35	.2236**	.0092	.0715	.0137	.2094**	.0629	.0333	-.0149	-.0351	-.0237	1.0000
ITEM36	.2106**	.0461	.0932	.0523	.2321**	.0579	.0256	.0490	-.0278	-.0357	.8744**
ITEM37	.0894	.0262	-.1389	-.0120	-.2350**	.1614*	.0727	.0102	.1603*	.0745	.1007
ITEM38	.1533*	.1303	.0777	.0579	.1778*	.1058	-.0445	-.0697	.1380	.0583	.2588**
ITEM39	.0772	.0092	.0240	.1057	.1514	.1642*	.2104**	.0313	.0580	.0881	.0755
ITEM40	.1586*	-.0913	.0132	.1215	.0925	.0425	.0560	.0617	.1105	.1710*	-.0002
ITEM41	.1313	.0179	-.1162	.1186	.0648	.0150	.0076	.0749	.1364	.0046	.1341
ITEM42	.1522	.0190	.0959	.1383	.0142	.0443	.0749	.1166	-.0411	.1520	.2041**
ITEM44	-.1003	.0542	-.1283	-.0749	-.0185	-.0719	.0508	-.0500	-.1081	.0140	.0092
ITEM45	.2236**	.0642	.1007	.2023**	.0626	.1365	.0626	.0917	.1150	.1327	.0549
ITEM46	.1039	.1707*	.1023	-.0163	.1820*	.1006	-.0700	.0092	.0872	.0407	.1820*
ITEM47	.0909	.1212	-.0100	.2220**	.1402	.1275	.0572	.0638	-.0101	.0179	.2571**
ITEM48	.1377	.0863	-.0249	.0865	.1251	.0132	-.0129	.0329	-.0525	-.0045	.1163

	ITEM24	ITEM25	ITEM26	ITEM28	ITEM29	ITEM30	ITEM31	ITEM32	ITEM33	ITEM34	ITEM35
ITEM49	-.0582	-.0180	-.0518	.0816	-.1856*	.0923	.0727	.0650	.0121	.0672	-.0101
ITEM50	.0073	.1795*	-.0337	.1336	.0421	.0157	-.1014	-.0129	.0173	-.0162	.0644
ITEM51	.1029	.1200	-.0718	.0619	.0786	.1006	-.1133	.0297	.1017	.0797	.0950
ITEM52	.2424**	.0345	-.0497	.1137	-.0349	.1301	.0202	.0574	.0455	-.0089	.0582
ITEM53	.1950*	.1977*	.1291	.1890*	.1723*	.1398	-.0752	.2625**	.1721*	.1110	.0934
ITEM54	.0983	-.0089	.0461	.2034**	-.0208	.1379	.0360	.2664**	.1407	.1948*	-.0966
ITEM55	.0913	.1032	.0704	.2026**	-.0280	.1177	.0320	.0271	.1788*	.0339	.0074
ITEM56	-.0168	.0514	-.0025	-.0683	.0658	.0419	.0658	.2094**	.0316	.1749*	.0264
ITEM57	.1478	-.0614	-.1038	.0956	.0748	.1599*	.1028	.0818	.0905	.0611	.1296
ITEM59	.0664	.1731*	-.0100	.0738	.1402	.0234	-.1089	.0889	-.0101	.0670	.0211
ITEM60	.1608*	.1871*	.0669	.1192	.0285	.0617	.0570	-.0846	.0318	.1334	.1545*
ITEM61	.0601	.1203	-.1091	.1118	.1502	.1309	.0403	.0465	-.0525	.0732	.1231
ITEM62	.1573*	.0817	-.0157	.1223	.1080	.0224	.0516	.1788*	.0866	.2291**	.0079
ITEM64	.0381	.1313	.0406	.2990**	.0320	.1177	-.1180	.0815	.1174	.0605	.0358
ITEM65	.0713	.0258	-.1165	.0005	-.1043	.0225	.0055	.0313	.0388	-.0842	.1314
ITEM66	.1851*	.0388	-.1231	.1128	.1777*	.0155	.0847	.0544	-.0296	.1651*	.0548
ITEM67	-.1034	-.0092	.0053	.0838	-.0039	.0209	-.1219	-.0045	-.0882	.0429	.1200
ITEM69	.0636	.0944	-.0396	.0224	.1939*	-.0863	.0293	.0836	.0385	-.0022	.3540**
ITEM70	.0667	-.0269	-.0930	-.0393	-.0458	.0666	-.0458	.0923	-.1001	-.0624	-.0365
ITEM71	-.0693	.0086	-.1683*	-.0721	-.0183	-.0469	.0914	-.0754	.0595	-.0155	.0338
ITEM72	.0450	.1200	.0866	.0427	-.0091	.2533**	-.0091	.0008	.0527	.0886	-.0079
ITEM73	.0479	.0444	-.0103	.0818	.0927	.0311	-.0776	-.0686	.0245	-.0825	.1185
ITEM74	.0362	.2212**	-.0587	.1052	.1000	.0716	-.0698	.1272	.1620*	.0548	.0827
ITEM75	.0489	.2149**	-.0472	.1643*	.0477	.1016	.0752	.0619	.0249	-.0377	.1670*
ITEM76	.0418	.1212	.0999	.2813**	.1679*	.0234	-.0258	.1140	.1598*	-.0067	.1260
ITEM77	.0374	.1890*	.1364	.1530*	.1245	.1103	-.0952	.1030	.1368	.0976	.0069
ITEM78	-.0392	.0601	-.0149	.0069	.0805	-.0513	-.0293	-.0089	-.0104	.0266	.1137
ITEM79	.0576	.0270	-.0156	.1060	-.0058	-.0105	-.0346	.0552	-.0247	.0023	-.0189
ITEM81	.0273	-.0093	-.1077	-.0348	-.0398	.1394	-.0398	-.1517	-.0786	.0072	.0582
ITEM82	.1048	-.1880*	-.1356	-.0853	-.0887	.0127	-.0887	.0144	-.0122	-.0624	.0042

	ITEM36	ITEM37	ITEM38	ITEM39	ITEM40	ITEM41	ITEM42	ITEM44	ITEM45	ITEM46	ITEM47
ITEM1	.2694**	-.0168	.2112**	.0256	.0170	.1925*	.2288**	-.1202	.0548	.0700	.1366
ITEM2	.0861	.0602	.1282	.0726	.1214	.0172	.0995	-.1435	.1273	-.0513	.0171
ITEM3	.0921	-.0092	-.0199	.0531	.0123	.0774	.1347	-.0841	-.0085	-.1086	.0724
ITEM4	.3008**	.0000	.1254	-.0567	.0732	.0382	.1142	-.0116	.0736	.1825*	.1226
ITEM5	.0415	-.0562	-.0018	.0663	.0160	.0217	.1480	-.1590*	.1585*	.0964	.0587
ITEM6	.1930*	.1093	-.0733	.0264	.1247	.1697*	.0349	-.1272	.0189	.0817	-.0379
ITEM7	.2325**	.0057	.0405	.0969	.1159	.1166	.1530*	-.1356	.1344	.0388	.1425
ITEM8	.1395	-.0105	.0938	.0273	-.0400	.0186	.2180**	-.1788*	.0372	.0612	.0721
ITEM9	.1385	.1160	.0371	.1395	.1433	.1161	.0682	-.2132**	.0734	-.1007	-.0199
ITEM10	.1568*	-.0909	-.1251	.0229	.0323	-.0296	.0011	.0301	.0742	-.0814	.0823
ITEM11	.0726	.1214	-.1031	.0437	.1870*	.0720	.2065**	-.0999	.1382	.0238	.0875
ITEM12	.2439**	.1603*	.1949*	.1184	.1105	.1656*	.2075**	-.0017	.1451	.0872	.1032
ITEM13	.0825	-.0599	.0251	-.1687*	-.0659	.0324	-.0688	.0779	-.0120	-.0335	.0446
ITEM14	.2471**	-.0823	.0260	.0627	.1745*	.0424	.1651*	.0229	.0680	.1149	.1351
ITEM16	.0750	-.0615	.1398	.0871	.1790*	.2046**	.0661	-.1959*	.2529**	.0233	.1584*
ITEM17	-.0283	.0105	.1943*	.0839	.2121**	.1431	-.0177	-.1806*	.1841*	.0576	.1627*
ITEM18	.1314	.0318	.1622*	.0362	.1731*	.1096	.2273**	-.1405	.1952*	.0566	.0575
ITEM19	.1583*	.2134**	.1040	-.0013	.2234**	.2328**	.0713	-.0388	.1715*	.0666	.1646*
ITEM20	.1673*	-.0105	.0152	-.0561	-.0056	-.0622	.0749	.0499	.0372	.0216	.0460
ITEM21	.2286**	.1447	.1761*	.0607	.0580	.0320	.1216	-.1534*	.1604*	.0647	.0745
ITEM22	.1231	.0067	-.0097	.0885	.0690	.1591*	.0977	-.1145	-.0938	-.0137	.0866
ITEM23	.0249	.0099	.0439	.0536	.1005	.2239**	.0364	-.0227	-.0674	.1569*	.0564
ITEM24	.2106**	.0894	.1533*	.0772	.1586*	.1313	.1522	-.1003	.2236**	.1039	.0909
ITEM25	.0461	.0262	.1303	.0092	-.0913	.0179	.0190	.0542	.0642	.1707*	.1212
ITEM26	.0932	-.1389	.0777	.0240	.0132	-.1162	.0959	-.1283	.1007	.1023	-.0100
ITEM28	.0523	-.0120	.0579	.1057	.1215	.1186	.1383	-.0749	.2023**	-.0163	.2220**
ITEM29	.2321**	-.2350**	.1778*	.1514	.0925	.0648	.0142	-.0185	.0626	.1820*	.1402
ITEM30	.0579	.1614*	.1058	.1642*	.0425	.0150	.0443	-.0719	.1365	.1006	.1275
ITEM31	.0256	.0727	-.0445	.2104**	.0560	.0076	.0749	.0508	.0626	-.0700	.0572
ITEM32	.0490	.0102	-.0697	.0313	.0617	.0749	.1166	-.0500	.0917	.0092	.0638
ITEM33	-.0278	.1603*	.1380	.0580	.1105	.1364	-.0411	-.1081	.1150	.0872	-.0101
ITEM34	-.0357	.0745	.0583	.0881	.1710*	.0046	.1520	.0140	.1327	.0407	.0179
ITEM35	.8744**	.1007	.2588**	.0755	-.0002	.1341	.2041**	.0092	.0549	.1820*	.2571**
ITEM36	1.0000	.0586	.2448**	.0671	-.0061	.1523	.2258**	-.0302	.0362	.1877*	.2734**
ITEM37	.0586	1.0000	.0602	-.0053	.1714*	.1704*	-.0877	.0250	.1007	.0379	.0350
ITEM38	.2448**	.0602	1.0000	-.1125	.0887	-.0084	.0995	-.0503	.1799*	.2498**	.0668
ITEM39	.0671	-.0053	-.1125	1.0000	.1104	.1471	.0343	-.1018	.1314	-.0279	.0164
ITEM40	-.0061	.1714*	.0887	.1104	1.0000	.1343	.0366	.0233	.2073**	.0195	.0962
ITEM41	.1523	.1704*	-.0084	.1471	.1343	1.0000	.0054	-.0141	.0799	-.0329	.1178
ITEM42	.2258**	-.0877	.0995	.0343	.0366	.0054	1.0000	-.2077**	.1466	.1521	.0210
ITEM44	-.0302	.0250	-.0503	-.1018	.0233	-.0141	-.2077**	1.0000	-.0237	-.0527	-.0857
ITEM45	.0362	.1007	.1799*	.1314	.2073**	.0799	.1466	-.0237	1.0000	.1820*	.1522
ITEM46	.1877*	.0379	.2498**	-.0279	.0195	-.0329	.1521	-.0527	.1820*	1.0000	.0227
ITEM47	.2734**	.0350	.0668	.0164	.0962	.1178	.0210	-.0857	.1522	.0227	1.0000
ITEM48	.0790	.0797	.1615*	.1312	.0199	.0738	.0867	-.0357	.2730**	.1065	.0884

	ITEM36	ITEM37	ITEM38	ITEM39	ITEM40	ITEM41	ITEM42	ITEM44	ITEM45	ITEM46	ITEM47
ITEM49	-.0744	.0311	-.1479	.0744	.0053	.1393	.0978	.0562	.0443	-.1155	-.2040**
ITEM50	.1017	-.1088	.0449	.0076	.1299	.0461	.2115**	-.1525	-.0443	.1544*	.0758
ITEM51	.0302	.1287	.1894*	-.0041	.0618	.0334	-.0441	-.0336	.2508**	.0765	.2763**
ITEM52	.0440	.1592*	.1249	.0871	.0506	-.0391	-.0249	.0108	.1104	.0385	.1930*
ITEM53	.0798	.0645	.1779*	-.0013	.0291	.1059	.1252	-.0695	.1976*	.1411	.0664
ITEM54	-.0601	-.0103	.0889	-.0480	.0452	.0737	.0383	-.2594**	.0109	.0618	.1331
ITEM55	.0445	.0650	.0675	.1268	.1525	.1528	-.0012	-.1509	.1494	.2057**	.1362
ITEM56	.0399	-.0396	-.0277	.0906	.0405	.1527	.2398**	-.0217	.0524	.0675	.1015
ITEM57	.1464	.1266	.1052	.0405	.1642*	.0455	.1232	-.1759*	.1030	.0334	.2262**
ITEM59	.0362	.0350	.0171	.0164	.0310	.2199**	-.0603	.0381	.1260	.0602	.2089**
ITEM60	.1109	.1800*	.0998	-.0296	.0732	.0645	.0863	.0840	.1545*	.0667	.1480
ITEM61	.1377	.0198	.1034	-.0331	.0626	.1386	.2440**	-.0901	.0971	.1613*	.3157**
ITEM62	.0244	-.0827	.0644	.1264	.1504	.0880	.1841*	-.0785	.1252	.0553	.0831
ITEM64	.0159	-.0162	.2019**	.0983	.0465	.0421	.1751*	-.0838	.1494	.0837	.1094
ITEM65	.1197	.1784*	.0398	-.0413	.0247	-.0732	-.0597	.0882	.1054	-.0203	.0107

ITEM66	.0924	-.0168	-.0111	.0551	.0901	.1068	.0162	-.0855	.2602**	.1120	.1919*
ITEM67	.0733	.0320	.0067	.1513	.0634	.1523	.1391	-.1292	.2597**	.0279	.1153
ITEM69	.3422**	-.0396	.1467	-.0288	-.0838	-.0950	.0942	.0405	.0682	.1273	.0979
ITEM70	-.0414	.0233	.0364	-.0404	-.0619	-.1503	.0247	-.0329	-.0771	-.0405	-.0067
ITEM71	-.0316	.0000	.0134	-.1772*	-.0460	-.0353	.1018	-.0149	-.0181	-.0259	.0594
ITEM72	.0055	-.0248	.2079**	.0728	.1106	.1365	-.0212	-.0465	.1739*	.2117**	.1856*
ITEM73	.0210	-.0359	.1398	.0331	.0452	.1261	-.0730	-.1642*	.0647	.1772*	.2852**
ITEM74	.1740*	.0562	.1287	-.0663	.0841	.0566	.1847*	-.1575*	.1095	.2103**	.1941*
ITEM75	.1819*	.1092	.0686	-.0249	-.0614	.0716	.1711*	-.1457	.1149	.0824	.2038**
ITEM76	.1153	-.1399	.0668	.0955	-.0342	.0667	.0210	-.1167	.0473	.0227	.2583**
ITEM77	.0715	.0546	.1675*	-.0453	.0669	.0642	.1866*	-.0942	.1110	.2109**	.1260
ITEM78	.2062**	.0149	.0747	-.0234	.1162	.0950	.0671	-.1019	.0617	.0957	.1227
ITEM79	-.0284	-.0052	-.0559	.0284	.0451	.0962	.1910*	-.0985	.1449	.0746	.0379
ITEM81	.0215	.1346	.1312	-.0499	.0532	.0584	.1098	.0435	.1147	.0024	.1461
ITEM82	-.0414	.0620	.0364	-.0404	.0392	.0082	-.1015	.0632	-.0365	-.0987	-.0834

	ITEM48	ITEM49	ITEM50	ITEM51	ITEM52	ITEM53	ITEM54	ITEM55	ITEM56	ITEM57	ITEM59
ITEM1	-.0147	.0421	.1301	.0311	-.0202	.1301	.1628*	.1180	.0713	.0374	.1089
ITEM2	.1862*	.1352	-.1095	.1157	.0013	.2026**	.2416**	.1481	.0952	.0046	.1412
ITEM3	.0340	-.0203	-.0740	-.0207	-.0828	.0908	.0556	.0069	.1072	.0516	.0724
ITEM4	-.0369	.0288	.0240	.0550	.0023	-.0414	-.1091	.0326	.1375	.0422	.1226
ITEM5	.0647	.1627*	.0732	.1204	.0958	-.0613	.0757	.1482	.1393	-.0109	.0587
ITEM6	.0569	.0010	.0257	.0171	-.0401	-.0321	.2093**	-.0259	.0760	.0901	-.0122
ITEM7	.0443	.0514	.0652	.0870	.1445	.1020	.0136	.0770	.1564*	.1010	.1144
ITEM8	.0548	.0275	-.0005	-.0042	-.1242	.0485	.0258	.1424	.1419	.0721	-.0583
ITEM9	.0752	-.0817	-.0266	.0817	.0203	.0551	.0545	-.0293	.0648	.0072	-.0199
ITEM10	.0397	-.1097	.0223	.0799	.0234	.0376	.0728	-.0465	-.0172	-.1010	.0542
ITEM11	.1537*	.0518	.0897	.0773	.0987	.1136	.0282	.1083	.0383	.0796	.0875
ITEM12	.1734*	.0121	.0760	.1297	.0173	.2283**	.0245	.1174	.1438	-.0244	.1315
ITEM13	.0870	.0244	.0595	.0573	.0932	.0506	-.0483	.0218	.0150	.0521	.0985
ITEM14	.1214	.1400	.0409	.1406	.0302	.1300	.0014	.0405	.0963	.0598	.1351
ITEM16	.1514	.0116	.0147	.2432**	.0667	.1487	.0377	.0518	.1242	.0265	.1838*
ITEM17	.2052**	-.0005	.0817	.2626**	.1242	.1847*	.0010	.2251**	.0390	-.0192	.1888*
ITEM18	.1973*	.0644	.1259	.1386	.0201	.1409	.1235	.1346	.1553*	.1892*	-.0998
ITEM19	.1132	-.0073	.1091	.0542	-.0999	.1950*	-.0024	.1445	.0562	-.1259	-.0073
ITEM20	.0288	-.0807	.0807	.0216	.0316	.0226	-.0812	-.1120	.0644	-.1394	-.0323
ITEM21	-.0702	-.0847	-.0177	.1169	.0809	.2002**	.0410	-.0433	-.0573	.0409	.0745
ITEM22	.0642	.0512	.0861	-.0137	-.0695	.0514	-.1012	.0711	-.0244	.0213	-.0457
ITEM23	-.0642	.0327	.1201	.0430	.0754	.0977	.1536*	.1482	.1627*	.0264	.1301
ITEM24	.1377	-.0582	.0073	.1029	.2424**	.1950*	.0983	.0913	-.0168	.1478	.0664
ITEM25	.0863	-.0180	.1795*	.1200	.0345	.1977*	-.0089	.1032	.0514	-.0614	.1731*
ITEM26	-.0249	-.0518	-.0337	-.0718	-.0497	.1291	.0461	.0704	-.0025	-.1038	-.0100
ITEM28	.0865	.0816	.1336	.0619	.1137	.1890*	.2034**	.2026**	-.0683	.0956	.0738
ITEM29	.1251	-.1856*	.0421	.0786	-.0349	.1723*	-.0208	-.0280	.0658	.0748	.1402
ITEM30	.0132	.0923	.0157	.1006	.1301	.1398	.1379	.1177	.0419	.1599*	.0234
ITEM31	-.0129	.0727	-.1014	-.1133	.0202	-.0752	.0360	.0320	.0658	.1028	-.1089
ITEM32	.0329	.0650	-.0129	.0297	.0574	.2625**	.2664**	.0271	.2094**	.0818	.0889
ITEM33	-.0525	.0121	.0173	.1017	.0455	.1721*	.1407	.1788*	.0316	.0905	-.0101
ITEM34	-.0045	.0672	-.0162	.0797	-.0089	.1110	.1948*	.0339	.1749*	.0611	.0670
ITEM35	.1163	-.0101	.0644	.0950	.0582	.0934	-.0966	.0074	.0264	.1296	.0211
ITEM36	.0790	-.0744	.1017	.0302	.0440	.0798	-.0601	.0445	.0399	.1464	.0362
ITEM37	.0797	.0311	-.1088	.1287	.1592*	.0645	-.0103	.0650	-.0396	.1266	.0350
ITEM38	.1615*	-.1479	.0449	.1894*	.1249	.1779*	.0889	.0675	-.0277	.1052	.0171
ITEM39	.1312	.0744	.0076	-.0041	.0871	-.0013	-.0480	.1268	.0906	.0405	.0164
ITEM40	.0199	.0053	.1299	.0618	.0506	.0291	.0452	.1525	.0405	.1642*	.0310
ITEM41	.0738	.1393	.0461	.0334	-.0391	.1059	.0737	.1528	.1527	.0455	.2199**
ITEM42	.0867	.0978	.2115**	-.0441	-.0249	.1252	.0383	-.0012	.2398**	.1232	-.0603
ITEM44	-.0357	.0562	-.1525	-.0336	.0108	-.0695	-.2594**	-.1509	-.0217	-.1759*	.0381
ITEM45	.2730**	.0443	-.0443	.2508**	.1104	.1976*	.0109	.1494	.0524	.1030	.1260
ITEM46	.1065	-.1155	.1544*	.0765	.0385	.1411	.0618	.2057**	.0675	.0334	.0602
ITEM47	.0884	-.2040**	.0758	.2763**	.1930*	.0664	.1331	.1362	.1015	.2262**	.2089**
ITEM48	1.0000	-.0680	.1702*	.1532*	-.0064	-.0092	.1008	.1784*	.1504	.1559*	.0391

	ITEM48	ITEM49	ITEM50	ITEM51	ITEM52	ITEM53	ITEM54	ITEM55	ITEM56	ITEM57	ITEM59
ITEM49	-.0680	1.0000	-.0694	-.0045	.0003	-.0837	-.0147	.0896	.0896	-.0220	-.0246
ITEM50	.1702*	-.0694	1.0000	-.0209	.0252	.0073	.1461	.0771	.0882	.1519	.0246
ITEM51	.1532*	-.0045	-.0209	1.0000	.4442**	.1758*	.1427	.0172	.1274	.1316	.3008**
ITEM52	-.0064	.0003	.0252	.4442**	1.0000	.1691*	.0415	-.0260	.0567	.2033**	.1930*
ITEM53	-.0092	-.0837	.0073	.1758*	.1691*	1.0000	.0479	.1445	.0075	.0234	.2628**
ITEM54	.1008	-.0147	.1461	.1427	.0415	.0479	1.0000	.2441**	.1493	.1806*	.1077
ITEM55	.1784*	.0896	.0771	.0172	-.0260	.1445	.2441**	1.0000	.1076	-.0189	.1094
ITEM56	.1504	.0896	.0882	.1274	.0567	.0075	.1493	.1076	1.0000	.0767	.2239**
ITEM57	.1559*	-.0220	.1519	.1316	.2033**	.0234	.1806*	-.0189	.0767	1.0000	.0008
ITEM59	.0391	-.0246	.0246	.3008**	.1930*	.2628**	.1077	.1094	.2239**	.0008	1.0000
ITEM60	.0646	.0288	.1032	.0802	.0023	.0850	-.0830	.1429	.0367	.0938	-.0301
ITEM61	.0717	-.0017	.0526	.1037	-.0114	.1089	.0369	.1560*	.2236**	.1160	.2666**
ITEM62	.0380	-.0327	.0852	.0903	-.0129	.0917	.0627	.1672*	.2598**	-.0605	.0326
ITEM64	.1249	-.0216	.0216	.1764*	-.0260	.1179	.0243	-.0450	-.0781	.0897	-.0245
ITEM65	-.0124	-.0106	-.0911	.0428	.1460	.0956	-.1047	.0309	-.1042	-.1174	.0352
ITEM66	.0681	.1282	-.0708	.1133	.0625	.0752	-.0360	.1480	.0713	.0654	-.0295
ITEM67	.2891**	.1169	.0744	.1346	-.0347	-.0249	-.0061	.1016	.1442	.0396	.0626
ITEM69	.1942*	-.2091**	.0567	.1393	.0627	.0879	-.0554	-.0122	.0432	.1023	.0489
ITEM70	-.1684*	-.1463	-.0923	-.0212	.0963	-.0856	.0403	-.0980	-.1264	-.0577	-.1218
ITEM71	-.0392	-.0477	.0477	.1029	.0273	-.1423	-.0396	-.0201	.0794	-.0617	.1084
ITEM72	-.0096	.0731	.1047	.0911	.1190	.0693	.0647	.2324**	.1389	-.0128	.1366
ITEM73	.1766*	-.0673	.1461	.1679*	.0415	.0227	.1677*	.0793	-.0014	.0779	.1077
ITEM74	.1620*	-.0054	.3199**	.1549*	.0300	.1115	.1576*	.1804*	.1611*	.1901*	.1941*
ITEM75	.1316	-.0182	.1455	.1403	.0754	.0733	.1536*	-.0115	.0654	.1757*	.1546*
ITEM76	.1377	-.1271	.1784*	.1049	.0453	.0909	.1584*	.1094	.0280	.1761*	.1594*
ITEM77	-.0228	-.1764*	.1764*	.1150	-.1107	.2323**	.1141	.1099	.1895*	-.0166	.0769
ITEM78	.0012	-.0704	.0704	.0791	-.0627	.0094	-.0200	-.0143	.1266	.0715	.0982
ITEM79	.0714	.0257	.0544	.0848	.0144	-.0191	.0547	.0259	.0515	.1446	.0379
ITEM81	.1561*	.0117	-.0670	.1543*	.0643	-.0521	-.0414	.1111	.1175	.2241**	-.0404
ITEM82	-.0920	-.0270	-.0525	-.0591	-.0946	-.0095	-.0777	.0267	-.0884	-.0966	-.1210

	ITEM60	ITEM61	ITEM62	ITEM64	ITEM65	ITEM66	ITEM67	ITEM69	ITEM70	ITEM71	ITEM72
ITEM1	.0570	.0421	.0330	.0280	-.0604	.0393	.0629	.0805	.0458	.1280	.0640
ITEM2	.0743	.2020**	.1247	.0944	.0398	.1000	.0596	-.0501	-.0406	-.0358	-.0380
ITEM3	.0128	.0932	.1422	.1053	.0967	.1289	-.0047	.1700*	.0350	.0314	-.1214
ITEM4	-.0214	.0505	.0627	.0602	-.0413	.0285	.0838	.1582*	.0359	.0733	.1535*
ITEM5	.0402	.0518	.0981	-.0161	-.0829	.0717	.2031**	.1174	-.1231	.2018**	.1487
ITEM6	.0072	.0664	.0852	.0857	-.1035	.0519	.1382	.1457	.0225	-.1182	-.0348
ITEM7	.1946*	.1283	.0971	.0465	-.0598	.1825*	.1427	.0527	-.0507	.0177	.0936
ITEM8	.1245	.2278**	.0352	.1424	-.1430	.0701	.0561	.1514	-.0898	-.0728	.0728
ITEM9	-.0708	.0052	.0285	.1120	-.0898	.0175	.0829	.0813	.0088	-.0564	-.0470
ITEM10	-.0210	-.0167	.0175	-.0161	.0319	-.0252	-.0828	.0309	.0071	-.1291	-.0101
ITEM11	.1328	.1770*	.1683*	.1674*	.0881	.0693	.1599*	.0818	-.0126	-.1270	.0459
ITEM12	.1484	.0600	.1342	.3015**	.0388	.1290	.0929	.1227	-.1001	.0315	-.0315
ITEM13	.0151	-.0695	.0201	.0510	-.0617	-.0543	-.0611	.0806	-.1452	-.0607	-.0194
ITEM14	.0245	.1520	.0347	.0405	.0587	.0369	.1232	.1497	-.0620	-.0742	-.0122
ITEM16	.0214	-.0638	.0725	.0793	-.0041	.1060	.2912**	-.0052	-.0777	.1111	.1903*
ITEM17	.1709*	.0311	.1083	.0838	-.0381	.1928*	.2775**	.0813	-.1530*	.1245	.1598*
ITEM18	.1153	.1110	.1264	.0778	-.0534	.1800*	.0755	.0098	.0771	-.0858	-.0700
ITEM19	.1356	.0845	.0917	.1977*	.1444	.1301	.0536	.0879	-.0856	-.0206	.0450
ITEM20	.1245	-.0570	-.0817	.0858	.0898	.0117	.0561	.2548**	.0316	.0822	.0211
ITEM21	.0808	.0833	.0479	.1439	-.0988	.0332	.0971	.0605	.0654	.0711	.0267
ITEM22	.1424	.0854	.0518	.0352	.0251	-.0074	.1231	.1664*	.0056	.0626	-.0298
ITEM23	-.0345	.0374	.1840*	.0684	-.0469	-.0202	.0772	-.0392	-.1048	-.0280	.0280
ITEM24	.1608*	.0601	.1573*	.0381	.0713	.1851*	-.1034	.0636	.0667	-.0693	.0450
ITEM25	.1871*	.1203	.0817	.1313	.0258	.0388	-.0092	.0944	-.0269	.0086	.1200
ITEM26	.0669	-.1091	-.0157	.0406	-.1165	-.1231	.0053	-.0396	-.0930	-.1683*	.0866
ITEM28	.1192	.1118	.1223	.2990**	.0005	.1128	.0838	.0224	-.0393	-.0721	.0427
ITEM29	.0285	.1502	.1080	.0320	-.1043	.1777*	-.0039	.1939*	-.0458	-.0183	-.0091
ITEM30	.0617	.1309	.0224	.1177	.0225	.0155	.0209	-.0863	.0666	-.0469	.2533**
ITEM31	.0570	.0403	.0516	-.1180	.0055	.0847	-.1219	.0293	-.0458	.0914	-.0091
ITEM32	-.0846	.0465	.1788*	.0815	.0313	.0544	-.0045	.0836	.0923	-.0754	.0008

ITEM33	.0318	-.0525	.0866	.1174	.0388	-.0296	-.0882	.0385	-.1001	.0595	.0527
ITEM34	.1334	.0732	.2291**	.0605	-.0842	.1651*	.0429	-.0022	-.0624	-.0155	.0886
ITEM35	.1545*	.1231	.0079	.0358	.1314	.0548	.1200	.3540**	-.0365	.0338	-.0079
ITEM36	.1109	.1377	.0244	.0159	.1197	.0924	.0733	.3422**	-.0414	-.0316	.0055
ITEM37	.1800*	.0198	-.0827	-.0162	.1784*	-.0168	.0320	-.0396	.0233	.0000	-.0248
ITEM38	.0998	.1034	.0644	.2019**	.0398	-.0111	.0067	.1467	.0364	.0134	.2079**
ITEM39	-.0296	-.0331	.1264	.0983	-.0413	.0551	.1513	-.0288	-.0404	-.1772*	.0728
ITEM40	.0732	.0626	.1504	.0465	.0247	.0901	.0634	-.0838	-.0619	-.0460	.1106
ITEM41	.0645	.1386	.0880	.0421	-.0732	.1068	.1523	-.0950	-.1503	-.0353	.1365
ITEM42	.0863	.2440**	.1841*	.1751*	-.0597	.0162	.1391	.0942	.0247	.1018	-.0212
ITEM44	.0840	-.0901	-.0785	-.0838	.0882	-.0855	-.1292	.0405	-.0329	-.0149	-.0465
ITEM45	.1545*	.0971	.1252	.1494	.1054	.2602**	.2597**	.0682	-.0771	-.0181	.1739*
ITEM46	.0667	.1613*	.0553	.0837	-.0203	.1120	.0279	.1273	-.0405	-.0259	.2117**
ITEM47	.1480	.3157**	.0831	.1094	.0107	.1919*	.1153	.0979	-.0067	.0594	.1856*
ITEM48	.0646	.0717	.0380	.1249	-.0124	.0681	.2891**	.1942*	-.1684*	-.0392	-.0053

	ITEM60	ITEM61	ITEM62	ITEM64	ITEM65	ITEM66	ITEM67	ITEM69	ITEM70	ITEM71	ITEM72
ITEM49	.0288	-.0017	-.0327	-.0216	-.0106	.1282	.1169	-.2091**	-.1463	-.0477	.0731
ITEM50	.1032	.0526	.0852	.0216	-.0911	-.0708	.0744	.0567	-.0923	.0477	.1047
ITEM51	.0802	.1037	.0903	.1764*	.0428	.1133	.1346	.1393	-.0212	.1029	.0911
ITEM52	.0023	-.0114	-.0129	-.0260	.1460	.0625	-.0347	.0627	.0963	.0273	.1190
ITEM53	.0045	.0850	.1089	.0917	.1179	.0956	.0752	-.0249	.0879	-.0856	-.1423
ITEM54	-.0830	.0369	.0627	.0243	-.1047	-.0360	-.0061	-.0554	.0403	-.0396	.0647
ITEM55	.1429	.1560*	.1672*	-.0450	.0309	.1480	.1016	-.0122	-.0980	-.0201	.2324**
ITEM56	.0367	.2236**	.2598**	-.0781	-.1042	.0713	.1442	.0432	-.1264	.0794	.1389
ITEM57	.0938	.1160	-.0605	.0897	-.1174	.0654	.0396	.1023	-.0577	-.0617	-.0128
ITEM59	-.0301	.2666**	.0326	-.0245	.0352	-.0295	.0626	.0489	-.1218	.1084	.1366
ITEM60	1.0000	.1515	.0627	.0602	-.0918	.2279**	.0838	.1330	-.0825	-.0275	.0779
ITEM61	.1515	1.0000	.2342**	.1028	-.0584	.1245	-.0192	.1913*	-.1091	.0081	.0891
ITEM62	.0627	.2342**	1.0000	.1300	-.0360	.2023**	.0616	.0110	.0229	-.0293	.0543
ITEM64	.0602	.1028	.1300	1.0000	-.1551*	.0880	.1301	-.0122	.0267	-.0201	.1793*
ITEM65	-.0918	-.0584	-.0360	-.1551*	1.0000	-.0878	-.0371	.0570	.0768	.0685	-.0442
ITEM66	.2279**	.1245	.2023**	.0880	-.0878	1.0000	.1219	.0530	.0887	.0457	.1188
ITEM67	.0838	-.0192	.0616	.1301	-.0371	.1219	1.0000	.1071	-.0414	.0728	.1622*
ITEM69	.1330	.1913*	.0110	-.0122	.0570	.0530	.1071	1.0000	-.0343	.1626*	.0074
ITEM70	-.0825	-.1091	.0229	.0267	.0768	.0887	-.0414	-.0343	1.0000	.0460	-.0081
ITEM71	-.0275	.0081	-.0293	-.0201	.0685	.0457	.0728	.1626*	.0460	1.0000	.0434
ITEM72	.0779	.0891	.0543	.1793*	-.0442	.1188	.1622*	.0074	-.0081	.0434	1.0000
ITEM73	.0997	.0621	.1243	.0243	-.0293	.0492	.1291	.0451	-.0777	.0609	.1401
ITEM74	.0899	.3495**	.0824	.0983	.0328	.0415	-.0145	.1331	-.1121	.0235	.1017
ITEM75	.0919	.2567**	.1088	.1216	-.0226	.0348	.0772	.1068	.0856	.1423	.0037
ITEM76	.0717	.1685*	-.0178	.0826	-.0138	-.0295	-.0164	.2204**	-.0834	.0349	-.0594
ITEM77	.2273**	.1201	.0917	.1099	-.1363	-.0147	-.0070	.1492	-.0812	-.0081	.1296
ITEM78	-.0573	.1978*	.1043	.0653	-.1299	.0841	.1018	.0686	.0723	.0074	.0169
ITEM79	.0988	.1124	-.0752	.0538	.0524	.0922	.1636*	.0074	.0174	.0672	.0093
ITEM81	.0598	.0705	-.0853	.1111	-.0432	.0398	.1351	-.0016	-.0313	.0592	.0992
ITEM82	.0754	-.1091	.0790	-.0980	-.0754	.0458	-.0005	-.0723	.1078	.0460	-.0483

	ITEM73	ITEM74	ITEM75	ITEM76	ITEM77	ITEM78	ITEM79	ITEM81	ITEM82
ITEM1	.0208	.1830*	.1998*	.0812	.1776*	.1390	.0346	.0696	.0029
ITEM2	.1398	.0018	.1425	.0916	.0197	.0747	.1250	-.0024	.0364
ITEM3	.0090	.0053	.0897	.2087**	.1323	.0551	.0705	-.1009	.0350
ITEM4	.0736	.0378	.1425	-.0046	.1263	.0940	-.0072	-.0498	.0359
ITEM5	.1276	.0077	.0362	.0081	-.0017	-.0172	.0416	.0602	-.0839
ITEM6	-.0547	.0943	.0065	-.0637	.1124	.0584	-.0619	-.0756	-.0574
ITEM7	.1577*	.0924	.1771*	.0301	.1785*	.1980*	.0442	-.0385	.0800
ITEM8	-.0010	.2211**	.1329	.0721	.1346	-.0997	.0909	.0179	.0316
ITEM9	.0277	.0456	.0485	.0583	.1502	-.0038	-.0365	-.1303	-.0721
ITEM10	-.0136	-.0350	.1578*	.0542	.0725	-.1423	-.1319	.0083	.0071
ITEM11	-.0278	.1754*	.0763	.0056	-.0415	.0535	.0315	.0312	-.0549
ITEM12	.0245	.2489**	.1656*	.1315	.3055**	.0177	.0638	.0740	-.0122
ITEM13	-.0483	-.0010	.0029	.1524	-.0642	.0529	-.0117	-.0422	.0220
ITEM14	-.0880	.0975	.0432	.1351	.0212	.0520	-.0039	.0152	-.1522

ITEM16	.2457**	.0539	.1284	.0317	.0638	-.0702	.0283	.0952	-.0384
ITEM17	.1883*	.0989	.0744	.1627*	.1243	-.0296	.1264	.0102	-.0721
ITEM18	.0109	.1049	.0934	.0837	.1492	.0942	.1280	.1677*	-.0042
ITEM19	.0227	.1617*	.0977	.0664	.1836*	.0581	.0576	-.0257	-.0476
ITEM20	.0525	.0344	.1070	.0199	.0828	.0555	-.0178	.0179	-.1302
ITEM21	.0156	.0266	.0941	-.0489	.1862*	.2086**	-.0224	-.0483	-.0494
ITEM22	.1024	-.0387	.0472	.0866	.0788	.1616*	.0802	-.0292	.0569
ITEM23	.0024	.1646*	.0242	.1301	.2064**	.0149	-.0321	-.1595*	.0095
ITEM24	.0479	.0362	.0489	.0418	.0374	-.0392	.0576	.0273	.1048
ITEM25	.0444	.2212**	.2149**	.1212	.1890*	.0601	.0270	-.0093	-.1880*
ITEM26	-.0103	-.0587	-.0472	.0999	.1364	-.0149	-.0156	-.1077	-.1356
ITEM28	.0818	.1052	.1643*	.2813**	.1530*	.0069	.1060	-.0348	-.0853
ITEM29	.0927	.1000	.0477	.1679*	.1245	.0805	-.0058	-.0398	-.0887
ITEM30	.0311	.0716	.1016	.0234	.1103	-.0513	-.0105	.1394	.0127
ITEM31	-.0776	-.0698	.0752	-.0258	-.0952	-.0293	-.0346	-.0398	-.0887
ITEM32	-.0686	.1272	.0619	.1140	.1030	-.0089	.0552	-.1517	.0144
ITEM33	.0245	.1620*	.0249	.1598*	.1368	-.0104	-.0247	-.0786	-.0122
ITEM34	-.0825	.0548	-.0377	-.0067	.0976	.0266	.0023	.0072	-.0624
ITEM35	.1185	.0827	.1670*	.1260	.0069	.1137	-.0189	.0582	.0042
ITEM36	.0210	.1740*	.1819*	.1153	.0715	.2062**	-.0284	.0215	-.0414
ITEM37	-.0359	.0562	.1092	-.1399	.0546	.0149	-.0052	.1346	.0620
ITEM38	.1398	.1287	.0686	.0668	.1675*	.0747	-.0559	.1312	.0364
ITEM39	.0331	-.0663	-.0249	.0955	-.0453	-.0234	.0284	-.0499	-.0404
ITEM40	.0452	.0841	-.0614	-.0342	.0669	.1162	.0451	.0532	.0392
ITEM41	.1261	.0566	.0716	.0667	.0642	.0950	.0962	.0584	.0082
ITEM42	-.0730	.1847*	.1711*	.0210	.1866*	.0671	.1910*	.1098	-.1015
ITEM44	-.1642*	-.1575*	-.1457	-.1167	-.0942	-.1019	-.0985	.0435	.0632
ITEM45	.0647	.1095	.1149	.0473	.1110	.0617	.1449	.1147	-.0365
ITEM46	.1772*	.2103**	.0824	.0227	.2109**	.0957	.0746	.0024	-.0987
ITEM47	.2852**	.1941*	.2038**	.2583**	.1260	.1227	.0379	.1461	-.0834
ITEM48	.1766*	.1620*	.1316	.1377	-.0228	.0012	.0714	.1561*	-.0920

	ITEM73	ITEM74	ITEM75	ITEM76	ITEM77	ITEM78	ITEM79	ITEM81	ITEM82
ITEM49	-.0673	-.0054	-.0182	-.1271	-.1764*	-.0704	.0257	.0117	-.0270
ITEM50	.1461	.3199**	.1455	.1784*	.1764*	.0704	.0544	-.0670	-.0525
ITEM51	.1679*	.1549*	.1403	.1049	.1150	.0791	.0848	.1543*	-.0591
ITEM52	.0415	.0300	.0754	.0453	-.1107	-.0627	.0144	.0643	-.0946
ITEM53	.0227	.1115	.0733	.0909	.2323**	.0094	-.0191	-.0521	-.0095
ITEM54	.1677*	.1576*	.1536*	.1584*	.1141	-.0200	.0547	-.0414	-.0777
ITEM55	.0793	.1804*	-.0115	.1094	.1099	-.0143	.0259	.1111	.0267
ITEM56	-.0014	.1611*	.0654	.0280	.1895*	.1266	.0515	.1175	-.0884
ITEM57	.0779	.1901*	.1757*	.1761*	-.0166	.0715	.1446	.2241**	-.0966
ITEM59	.1077	.1941*	.1546*	.1594*	.0769	.0982	.0379	-.0404	-.1218
ITEM60	.0997	.0899	.0919	.0717	.2273**	-.0573	.0988	.0598	.0754
ITEM61	.0621	.3495**	.2567**	.1685*	.1201	.1978*	.1124	.0705	-.1091
ITEM62	.1243	.0824	.1088	-.0178	.0917	.1043	-.0752	-.0853	.0790
ITEM64	.0243	.0983	.1216	.0826	.1099	.0653	.0538	.1111	-.0980
ITEM65	-.0293	.0328	-.0226	-.0138	-.1363	-.1299	.0524	-.0432	-.0754
ITEM66	.0492	.0415	.0348	-.0295	-.0147	.0841	.0922	.0398	.0458
ITEM67	.1291	-.0145	.0772	-.0164	-.0070	.1018	.1636*	.1351	-.0005
ITEM69	.0451	.1331	.1068	.2204**	.1492	.0686	.0074	-.0016	-.0723
ITEM70	-.0777	-.1121	.0856	-.0834	-.0812	.0723	.0174	-.0313	.1078
ITEM71	.0609	.0235	.1423	.0349	-.0081	.0074	.0672	.0592	.0460
ITEM72	.1401	.1017	.0037	-.0594	.1296	.0169	.0093	.0992	-.0460
ITEM73	1.0000	.0539	.1536*	.1331	.0134	.0303	.0811	-.0141	.0010
ITEM74	.0539	1.0000	.3153**	.1688*	.2274**	.1174	.1426	-.0058	-.1121
ITEM75	.1536*	.3153**	1.0000	.3020**	.2064**	.1123	.0958	.0786	-.0286
ITEM76	.1331	.1688*	.3020**	1.0000	.1505	.0247	.2696**	.0662	-.1218
ITEM77	.0134	.2274**	.2064**	.1505	1.0000	.1670*	.0920	-.0969	-.1192
ITEM78	.0303	.1174	.1123	.0247	.1670*	1.0000	.0691	.0016	-.0037
ITEM79	.0811	.1426	.0958	.2696**	.0920	.0691	1.0000	.1034	-.0225
ITEM81	-.0141	-.0058	.0786	.0662	-.0969	.0016	.1034	1.0000	.1340
ITEM82	.0010	-.1121	-.0286	-.1218	-.1192	-.0037	-.0225	.1340	1.0000

APPENDIX D

Item parameter estimates for statistical knowledge items

ITEM	INTER S.E.	SLOPE S.E.	THRESH S.E.	DISPER S.E.	ASYMP S.E.	CHISQ (PROB)	DF
0001	0.495	0.822	-0.602	1.216	0.192	2.9	4.0
	0.169*	0.179*	0.255*	0.265*	0.084*	(0.5795)	
0002	-0.099	0.703	0.141	1.423	0.201	5.7	5.0
	0.201*	0.191*	0.273*	0.387*	0.084*	(0.3373)	
0003	1.575	0.862	-1.827	1.160	0.205	1.9	3.0
	0.255*	0.255*	0.442*	0.342*	0.091*	(0.6008)	
0004	0.063	0.674	-0.093	1.484	0.217	0.9	5.0
	0.194*	0.173*	0.298*	0.381*	0.089*	(0.9691)	
0005	-1.132	0.639	1.772	1.566	0.246	5.4	6.0
	0.440*	0.223*	0.534*	0.546*	0.069*	(0.4947)	
0006	0.054	0.551	-0.098	1.816	0.253	6.9	5.0
	0.208*	0.148*	0.386*	0.488*	0.099*	(0.2241)	
0007	0.536	0.925	-0.580	1.082	0.208	2.2	4.0
	0.179*	0.238*	0.245*	0.278*	0.089*	(0.7047)	
0008	-1.125	0.821	1.370	1.217	0.179	3.7	6.0
	0.386*	0.272*	0.328*	0.403*	0.059*	(0.7199)	
0009	0.167	0.696	-0.240	1.436	0.231	2.1	5.0
	0.194*	0.184*	0.302*	0.380*	0.094*	(0.8397)	
0010	-2.358	0.904	2.609	1.106	0.222	6.8	6.0
	0.946*	0.403*	0.826*	0.494*	0.042*	(0.3371)	
0011	-1.558	1.170	1.331	0.854	0.164	3.0	5.0
	0.552*	0.458*	0.280*	0.334*	0.050*	(0.7076)	
0012	0.749	1.297	-0.577	0.771	0.162	5.5	3.0
	0.184*	0.320*	0.175*	0.190*	0.074*	(0.1370)	
0013	0.309	0.439	-0.705	2.278	0.218	8.3	5.0

	0.167*	0.117*	0.451*	0.608*	0.094*	(0.1407)
0014	-1.801	1.126	1.600	0.888	0.148	11.1 5.0
	0.611*	0.430*	0.326*	0.339*	0.044*	(0.0499)
0015	0.062	0.849	-0.073	1.178	0.206	3.5 5.0
	0.196*	0.220*	0.239*	0.305*	0.084*	(0.6296)
0016	0.285	0.974	-0.293	1.026	0.183	3.9 5.0
	0.175*	0.244*	0.207*	0.257*	0.079*	(0.5655)
0017	-1.298	1.111	1.168	0.900	0.176	3.8 6.0
	0.454*	0.371*	0.234*	0.301*	0.053*	(0.7016)
0018	-0.164	0.729	0.225	1.373	0.164	7.9 5.0
	0.187*	0.184*	0.237*	0.347*	0.072*	(0.1592)
0019	-1.167	0.578	2.020	1.731	0.206	11.7 7.0
	0.413*	0.198*	0.607*	0.592*	0.066*	(0.1107)
0020	-0.533	0.669	0.797	1.494	0.168	13.1 6.0
	0.240*	0.163*	0.284*	0.364*	0.069*	(0.0406)
0021	0.892	0.644	-1.384	1.552	0.206	3.8 4.0
	0.171*	0.158*	0.401*	0.381*	0.091*	(0.4408)
0022	-1.264	1.048	1.206	0.954	0.324	5.5 6.0
	0.556*	0.413*	0.324*	0.376*	0.064*	(0.4798)
0023	-0.416	1.027	0.405	0.974	0.239	4.0 5.0
	0.288*	0.340*	0.218*	0.322*	0.080*	(0.5496)
0024	0.195	0.726	-0.269	1.377	0.207	2.4 5.0
	0.184*	0.159*	0.276*	0.301*	0.087*	(0.7898)
0025	0.725	0.921	-0.787	1.086	0.196	1.9 4.0
	0.176*	0.219*	0.253*	0.258*	0.086*	(0.7651)
0026	-1.305	0.907	1.439	1.103	0.138	4.0 5.0
	0.395*	0.279*	0.283*	0.339*	0.049*	(0.5458)
0027	-2.198	0.891	2.466	1.122	0.114	4.7 5.0
	0.710*	0.368*	0.673*	0.463*	0.036*	(0.4601)
0028	-2.661	0.912	2.916	1.096	0.243	2.5 6.0

	1.178*	0.425*	1.115*	0.510*	0.040*	(0.8745)	
0029	-0.148	0.657	0.225	1.522	0.269	8.8	5.0
	0.244*	0.192*	0.346*	0.445*	0.096*	(0.1177)	
0030	0.490	0.568	-0.863	1.760	0.211	6.5	5.0
	0.165*	0.143*	0.375*	0.444*	0.092*	(0.2604)	
0031	-1.317	1.131	1.164	0.884	0.318	8.1	6.0
	0.579*	0.447*	0.296*	0.349*	0.062*	(0.2335)	
0032	-0.065	1.686	0.038	0.593	0.322	7.8	4.0
	0.307*	0.606*	0.175*	0.213*	0.082*	(0.0993)	
0033	-0.023	1.866	0.012	0.536	0.319	6.5	4.0
	0.309*	0.707*	0.163*	0.203*	0.080*	(0.1617)	
0034	-0.101	0.434	0.233	2.304	0.239	11.4	6.0
	0.211*	0.121*	0.469*	0.640*	0.096*	(0.0762)	
0035	-0.086	0.719	0.120	1.390	0.195	4.1	5.0
	0.199*	0.175*	0.266*	0.338*	0.082*	(0.5309)	
0036	-1.610	0.799	2.015	1.252	0.235	2.1	7.0
	0.600*	0.312*	0.541*	0.490*	0.055*	(0.9545)	
0037	-2.654	1.577	1.683	0.634	0.124	2.8	5.0
	1.121*	0.766*	0.295*	0.308*	0.032*	(0.7331)	
0038	-1.053	0.887	1.186	1.127	0.191	4.2	6.0
	0.390*	0.302*	0.300*	0.383*	0.064*	(0.6494)	
0039	-1.296	1.008	1.285	0.992	0.146	3.8	5.0
	0.420*	0.329*	0.267*	0.324*	0.052*	(0.5785)	
0040	0.234	0.842	-0.278	1.187	0.213	1.6	5.0
	0.187*	0.220*	0.250*	0.310*	0.088*	(0.8979)	
0041	1.168	0.776	-1.505	1.289	0.218	1.2	2.0
	0.211*	0.234*	0.403*	0.388*	0.095*	(0.5556)	
0042	-0.804	1.196	0.672	0.836	0.192	11.6	5.0
	0.348*	0.407*	0.180*	0.284*	0.063*	(0.0397)	
0043	-0.741	0.876	0.846	1.141	0.219	2.8	6.0

	0.332*	0.293*	0.274*	0.382*	0.074*	(0.8326)	
0044	0.004	0.639	-0.007	1.565	0.264	16.4	5.0
	0.220*	0.192*	0.345*	0.470*	0.099*	(0.0060)	
0045	-0.523	0.957	0.547	1.045	0.228	1.4	5.0
	0.295*	0.314*	0.232*	0.343*	0.077*	(0.9224)	
0046	-0.677	0.785	0.862	1.273	0.341	8.2	6.0
	0.381*	0.288*	0.371*	0.467*	0.084*	(0.2224)	
0047	-0.243	0.807	0.301	1.239	0.193	2.7	5.0
	0.220*	0.222*	0.239*	0.341*	0.078*	(0.7429)	
0048	-0.025	0.561	0.044	1.782	0.249	5.3	5.0
	0.213*	0.158*	0.376*	0.503*	0.097*	(0.3752)	
0049	0.314	0.674	-0.466	1.483	0.213	7.3	5.0
	0.175*	0.174*	0.303*	0.383*	0.091*	(0.1982)	
0050	-0.694	0.851	0.815	1.175	0.267	3.9	6.0
	0.350*	0.298*	0.302*	0.412*	0.079*	(0.6881)	
0051	-1.261	1.033	1.220	0.968	0.254	3.1	6.0
	0.506*	0.391*	0.296*	0.366*	0.061*	(0.7985)	
0052	-0.815	0.861	0.946	1.161	0.224	4.0	6.0
	0.348*	0.277*	0.283*	0.374*	0.071*	(0.6741)	
0053	0.027	0.719	-0.038	1.391	0.233	1.6	5.0
	0.208*	0.194*	0.293*	0.375*	0.092*	(0.9063)	
0054	-0.856	1.235	0.693	0.810	0.244	3.7	5.0
	0.397*	0.431*	0.194*	0.283*	0.066*	(0.6004)	
0055	-1.629	1.525	1.068	0.656	0.257	4.2	6.0
	0.726*	0.684*	0.217*	0.294*	0.053*	(0.6566)	
0056	0.231	0.757	-0.306	1.320	0.258	4.2	5.0
	0.205*	0.204*	0.302*	0.356*	0.099*	(0.5180)	
0057	-2.576	0.952	2.707	1.051	0.428	4.6	7.0
	1.383*	0.455*	1.247*	0.503*	0.045*	(0.7128)	
0058	0.431	0.607	-0.710	1.646	0.208	7.2	5.0

	0.167*	0.143*	0.339*	0.388*	0.090*	(0.2040)	
0059	0.203	0.646	-0.314	1.548	0.235	4.5	5.0
	0.190*	0.172*	0.326*	0.412*	0.095*	(0.4766)	
0060	-0.280	0.655	0.428	1.528	0.190	1.1	5.0
	0.216*	0.159*	0.290*	0.372*	0.079*	(0.9512)	
0061	-0.636	0.442	1.441	2.264	0.312	8.3	6.0
	0.352*	0.139*	0.696*	0.712*	0.093*	(0.2137)	
0062	-0.573	0.605	0.947	1.654	0.246	8.2	6.0
	0.299*	0.188*	0.404*	0.515*	0.085*	(0.2250)	
0063	-0.087	0.682	0.127	1.466	0.272	5.0	5.0
	0.237*	0.205*	0.335*	0.441*	0.098*	(0.4138)	
0064	0.010	0.944	-0.011	1.060	0.218	7.4	5.0
	0.210*	0.270*	0.224*	0.303*	0.085*	(0.1902)	
0065	-0.620	0.992	0.625	1.008	0.199	5.2	5.0
	0.297*	0.319*	0.217*	0.324*	0.071*	(0.3929)	
0066	-0.742	0.873	0.850	1.145	0.205	1.8	6.0
	0.321*	0.282*	0.261*	0.369*	0.071*	(0.9358)	
0067	-0.179	0.723	0.247	1.382	0.161	3.4	5.0
	0.188*	0.171*	0.238*	0.327*	0.071*	(0.6395)	
0068	-0.719	0.668	1.076	1.497	0.274	10.2	6.0
	0.349*	0.224*	0.408*	0.501*	0.082*	(0.1170)	
0069	-1.379	0.818	1.685	1.222	0.238	6.7	6.0
	0.520*	0.301*	0.435*	0.450*	0.060*	(0.3469)	
0070	-2.140	0.842	2.541	1.187	0.260	5.8	7.0
	0.876*	0.365*	0.818*	0.515*	0.047*	(0.5613)	