Latent Trait, Factor, and Number Endorsed Scoring of Polychotomous and Dichotomous Responses to the Common Metric Questionnaire

by

R. Lance Becker

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

APPROVED:

Joseph J. Franchina, Chairman

Robert B. Frary

Roseanne J. Foti

Joseph A. Sgro

George A. Clum

December, 1991
Blacksburg, Virginia

Latent Trait, Factor, and Number Endorsed Scoring of Polychotomous and Dichotomous Responses to the Common Metric Questionnaire

by

R. Lance Becker

Committee Chair: Joseph J. Franchina

Psychology

(ABSTRACT)

Although job analysis is basic to almost all human resource functions, little attention has been given to the response format and scoring strategy of job analysis instruments. This study investigated three approaches to scoring polychotomous and dichotomous responses from the frequency and importance scales of the Common Metric Questionnaire (CMQ). Factor, latent trait, and number endorsed scores were estimated from the responses of 2684 job incumbents in six organizations. Scores from four of the CMQ scales were used in linear and nonlinear multiple regression equations to predict pay. The results demonstrated that: (a) simple number endorsed scoring of dichotomous responses was superior to the other scoring strategies; (b) Scoring of dichotomous responses was superior to scoring of polychotomous responses for each scoring technique; (c) scores estimated from the importance scale were better predictors of pay then scores from the frequency scale; (d) the relationship between latent trait and factor scores is nonlinear; (e) latent trait scores estimated with the two-parameter logistic model were superior to latent trait scores from the three-parameter model; (f) test information functions for each scale demonstrated that the CMQ scales accurately measured a relatively narrow range of theta; (g) the reliability of factor scores estimated from dichotomous data is superior to factor scores from polychotomous data. Issues regarding the construction of job analysis instruments and the use of item response theory are discussed.

ACKNOWLEDGEMENTS

I am grateful for the guidance I received from the members of my dissertation committee: Dr. Joseph J. Franchina, Dr. Robert B. Frary, Dr. Roseanne J. Foti, Dr. George A. Clum, and Dr. Joseph A. Sgro. I also appreciate the help and support I received from Dr. R. J. Harvey who provided the impetus for this project. I am especially grateful to Pamela whose love, support, and determination helped us through a very difficult period in our lives.

The Psychological Corporation requested that I include this statement:

"All Common Metric Questionnaire (CMQ) data is owned by The Psychological Corporation. The data was obtained for research purposes only and may therefore be unreliable. Researchers interested in obtaining copies of the CMQ data should contact The Psychological Corporation, 555 Academic Court, San Antonio, Texas 78204."

TABLE OF CONTENTS

LIST OF TABLES

Latent Trait, Factor, and Number Endorsed Scoring of Polychotomous and Dichotomous Responses to the Common Metric Questionnaire

One role of measurement in psychology is to quantify the nature and extent of individual differences or to assess the change that occurs in the same individual over time (Anastasi, 1976). Psychological tests are tools that are used to investigate these differences. Tests sample behavior in order to make accurate estimates concerning a broader repertoire. When the administration, scoring, and interpretation of the subject's responses are objective, tests may yield precise estimates of the true nature and extent of individual differences.

Lord and Novick (1968) defined measurement as a procedure for the assignment of numbers to the specific attributes or behaviors of experimental units in order to characterize and describe certain relationships. Measurement involves three basic processes: First, experimental units (ie,. individuals, organizations, positions) must be identified. Second, the attributes or behaviors of interest must be targeted. Third, according to a set of rules, an appropriate abstract number system (scale) must be used to assign numbers to the targeted attributes or behaviors. This procedure must produce data, which when subjected to mathematical operations, serves to organize, simplify, and describe the variables of interest. It is essential that these operations preserve the distinctiveness of the scores so that the extent and nature of the individual differences can be accurately identified. If the measurement system blurs these differences, valuable information about the attributes or behavior will be lost. Therefore, in order to capture accurate differences between individuals, great care must be exercised in the choice of the abstract number system, the arbitrary assignment of numbers to variables, and the mathematical operations used to analyze the data.

Psychological Tests in Employment

The use of psychological tests in employment situations is widespread (Hulin et al., 1983). Organizational functions such as selection, job specification, training and development, performance appraisal, and job evaluation use various tests to assess individual differences. These measures are used both as predictors and criteria. However before any of the above organizational functions can be initiated, it is necessary to gather information about the positions of interest (Gatewood & Field, 1987). Further, the Uniform Guidelines on Employee Selection Procedures (EEOC) and various litigation regarding discrimination in selection have supported the importance of job analysis (eg., Albemarle Paper Co. v. Moody, 1975; Griggs v. Duke Power Co., 1971).

According to Harvey (1991) job analysis is the gathering of data describing (a) the observable or verifiable worker behaviors, and (b) the objective attributes of the job environment which interact with the worker behaviors. Job analysis, as with any measurement approach, starts with defining the jobs that will be studied and continues along the same approach as outlined by Lord and Novick (1968.

Approaches to Job Analysis

There are basically two approaches to job analysis, task-oriented and worker-oriented. Task-oriented job analysis targets tasks that are basically well-defined purposeful work activities (Cascio, 1987). Although there have been many definitions for tasks, Harvey (1991) has found five similarities among the definitions: (a) Task statements involve action verbs or a series of actions. (b) Tasks are completed in order without long delays. (c) Tasks have discrete starting and stopping points. (d) Tasks have at least one identifiable purpose. (e) Tasks can be identified with individual positions. Task-oriented job analysis methods are used for organizational functions such as training and development, content validation strategies, and performance appraisals. However, because tasks are organization and technology specific, comparison of different jobs across these areas is not feasible. If the goal of the organization is to compare different jobs or to group jobs into job families based on similar work behaviors, worker-oriented items would be more appropriate than task-oriented items. Harvey (1991) notes that the difference between task-oriented and worker-oriented approaches to job analysis is largely illusory. However, they are used for different purposes. If the goal of the organization is to develop performance appraisal instruments, conduct content validation studies, or identify training or development goals, task-oriented job analysis will yield more useful information. If the organization's goal involves job evaluation, job classification, or clustering, then worker-oriented methods would be more informative. This approach provides cross-job comparisons.

Worker-oriented job analysis describes how the job is performed and concentrates on generalized work behaviors that are to some degree comparable across jobs (Cascio, 1987). According to McCormick, Jeanneret, and Mecham (1972) common denominators can be found to link jobs from different technologies together. These researchers assert that butchering, baking and candlestick making may share common work dimensions. Harvey (1991) points out that jobs that appear to be very different, when described in terms of the tasks, may have very similar work dimensions. According to Arvey and Begalla (1975), the work of a police officer and a housewife are very similar when compared on the basis of worker-oriented dimensions. Both jobs require resolving emergencies, enforcing rules, and maintaining order. Generalized work behaviors can be used to describe any job. However, job dimensions are not directly observable. The items which comprise the job analysis instrument are observable indicators that are used to identify and quantify the underlying dimensions. Item responses are scored in order to compute estimates of the unobservable work dimensions. Therefore, abstract work dimensions are constructs which must be inferred from the incumbent, supervisor, or observer's item responses.

The Position Analysis Questionnaire (McCormick et al., 1972) was designed to describe jobs in terms of worker-oriented job dimensions. The PAQ consists of 194 items that concern information input, cognitive processes (e.g. decision making), work output, interpersonal relationships (e.g. negotiating), and job context (e.g. working in adverse conditions). Reportedly, the PAQ measures 32 specific and 13 overall dimensions. Jobs are described in terms of how much of each dimension they possess. The PAQ is completed by either job analysts or job incumbents and supervisors who have received special training. Each position is scored on the work dimensions.

Although the authors of the PAQ assert that the instrument is capable of assessing the common denominators that are relevant to all jobs, Cascio (1987) identified two problems with the instrument. Studies of the usefulness of the PAQ suggest that it is more applicable to blue-collar manufacturing jobs than to managerial, technical or professional positions. Apparently, the PAQ has difficulty measuring abstract work dimensions associated with some higher functioning jobs. Another problem with the PAQ is that a college graduate reading level is necessary to use the instrument. Therefore, the PAQ can not be used by many incumbents or supervisors. Because of these limitations, the PAQ may not be applicable to certain kinds of jobs.

Although the PAQ appears to have some limitations, the advantages of a common work dimensions approach to job analysis continues to be voiced by other researchers. McCormick et al., (1972) and Harvey (1991) assert that worker-oriented job analysis can identify compensable factors. Compensable factors are the common work dimensions that an organization values and on which compensation levels are based. Job analysts have successfully used dimensions scores to account for a significant amount of the variance in pay. This process is termed a "policy capturing" approach to job evaluation.

Job Evaluation

Job evaluation is a process which management uses to set wage rates for each job in the organization. This process establishes the worth of jobs to the organization (Wallace & Fay, 1988). Job evaluation classifies jobs into a hierarchy that is used to set an equitable level of pay for each job. Job families are groupings of jobs that have a common set of characteristics. The purpose is to ensure that jobs of comparable worth receive comparable pay.

Although there are many approaches to setting wages, job evaluation based on information from the worker-oriented job analysis instruments are very useful (McCormick et al., 1972; and Harvey, in press). The four basic steps of job evaluation are: (a) Identify the jobs of interest. This decision will be based on the goals of the organization and reason for conducting a job evaluation. (b) Conduct a worker-oriented job analysis. This involves analyzing the responses of workers in order to identify and quantify the work dimensions associated with each position. (c) Use the work dimension scores to predict compensation rates (i.e., use computed regression coefficients to weight dimension scores in order to predict pay for each job). This method captures the pay policy of the organization and sets pay levels for jobs. According to some researchers this method has proven very productive: McCormick et al., (1977) reported multiple R for predicting pay from the PAQ dimension scores was .85. Harvey (1989) cited multiple Rs in the .70s and .80s for predicting pay from dimension scores as measured by the Job Element Inventory (JEI).

Scales and Response Formats

Type of scale and response format are two aspects of job analysis instruments that have not received much attention. It is plausible that the use of different types of scales (e.g., frequency, relative time spent) and response formats (eg., dichotomous, polychotomous) may affect the results of the job analysis.

Task-oriented inventories often rate each item on a number of scales. CODAP is a task-oriented job analysis system used by the United States Air Force (Christal, 1974). The questionnaire is sent by mail to workers who complete the inventory by simply checking whether they perform the task or not and estimating the amount of time, relative to other tasks, they spend performing the task. The incumbent is also directed to add task statements that are not listed to the questionnaire. The data regarding the performance of the task is binary, if the task is performed is receives a 1 if it is not performed it receives a 0. Data from the relative time spent scale is polychotomous. The relative time spent scale was incorporated in the job analysis instrument because it was believed that workers could not make accurate estimates of the absolute amount of time they devote to a task. However, they were hypothesized to be able to state whether they spend more or less time on one task relative to the other tasks they perform. The relative time spent scale ranges from "does not perform task" to "spends an extremely large amount of time on it compared to other tasks".

Functional Job Analysis (Fine & Wiley, 1971) uses Data, People, and Things functional scales, a worker instructions scale, reasoning development scale, a mathematical development scale, and a language development scale to measure different aspects of certain tasks. These scales consist of six- or eight-choice graded response categories.

The PAQ employs several graded response scales with either five- or seven-choice formats. Scales such as amount of time, noise intensity, degree of precision, level decision, level of reasoning in problem solving, importance, and extent of use are presented in different sections of the questionnaire. For example: The section titled "Use of Stored Information" asks for the "level of education generally or typically required by persons entering the occupation". The response format for this question is a likert-type scale with seven anchors: "Little or no formal education", "Elementary school (through sixth grade)", "Some high school (but not diploma)", "High school diploma", "Beyond high school (but not degree"), "College degree", and "Advanced degree (M.S., Ph.D., M.D., etc.)". Another scale used in many sections of the PAQ is the relative importance scale. It has a six-choice graded response format with the following anchors: "Does not apply", "Very minor", "Low", "Average", "High", and "Extreme". These responses are coded zero through five and the item level data are intercorrelated.

The Common Metric Questionnaire (CMQ) (Harvey, in press) is a worker-oriented job analysis questionnaire that is designed to characterize all jobs in all situations. Reportedly, the instrument has overcome the disadvantages and limitations of the PAQ: It can be completed by job incumbents and supervisors who have only a 9th grade reading level, and its underlying dimensions are relevant not only to blue-collar positions but also to professional, managerial, and scientific jobs.

The response format of the CMQ includes two five-choice graded response formats to assess frequency and importance. The anchors for the frequency scale are: "Hourly to many times an hour", "Daily", "Every few days to weekly", "Monthly", and "Every six months to yearly". The anchors for the relative importance scale are: "Very minor importance", "Below average importance", "Average importance", "Above average importance", and "Very high importance". If a worker-oriented question is not applicable to the position it receives a zero. If it does apply it receives a value of one through five.

Psychometric Issues

The approach used to compute factor scores for the CMQ is similar to the scoring strategy of the PAQ. However, the CMQ uses common factor analysis while the PAQ uses principal component analysis. Principal components analysis is a technique which uses weighted composites of the observed variables to summarize the variables (Cliff, 1987). Principal components analysis makes the assumptions that the observables are only influenced by the common factors and that they are measured without error. Also, principal components analysis assumes that the underlying dimensions are uncorrelated. The assumptions for factor analysis are quite different.

Factor analysis is concerned with explaining the variation in the manifest variables by identifying the underlying dimensions (Cliff, 1987). These dimensions are called common factors. Factor analytic techniques explain the regular patterns of the relationships among the observables. They are used for a variety of purposes such as defining the linear relationships among intercorrelated variables, identifying the basic structure of a domain, grouping variables into independent categories thus forming an empirical typology, data reduction by reducing large amounts of data to parsimonious factor patterns, hypothesis testing, exploration of a new domain in order to discover unknown relationships, and to aid in the building of models and the theories (Rummel, 1970). Factor analysis as a mathematical technique can be used in inductive or deductive theory building. In inductive theory building, it is used to discover important relationships in vast amounts of independent observations. This is a process in which the existence of constructs are inferred from empirical observations. In deductive theory building/testing, factor analysis is used to test for the existence of hypothesized underlying dimensions. After substantively motivated (rationally derived) assumptions are made about the interrelationships of the observations, factor analysis is can be used to test the validity of these assumptions.

There are basically two purposes for using factor analysis to analyze information from worker-oriented job analysis instruments. The first purpose is to identify the underlying dimensions of work. The second is to compute factor score estimates for each position. These are estimates of how much of the each factor is involved in each position. The procedure involves obtaining responses from a large number of positions to worker-oriented questions. The questionnaire is designed to investigate all aspects of worker behavior, products and environment. Therefore, the data consist of responses from a large number of positions to a smaller but still large number of questions. The responses to the items are correlated with one another and the intercorrelations are analyzed to identify the factors that could cause the variations in the data. According to Rummel (1970), cause refers to a uniform and functional relationship between the variables that is an expression of the concomitance of the phenomena. The analysis can yield one or more hypothetical causes that maybe useful to explain and classify the responses. Items that are influenced by the same factor will be grouped together. A rational interpretation of these items can reveal the identity of the factor. In situations where there are more than one substantively meaningful factor, items can be influenced by two or more factors. The amount of influence a factor has on each item is expressed in terms of the item's loading on that factor.

After the important work dimensions have been identified and loadings for each item have been computed, factor scores for each position can be generated. Therefore, each position will have scores on the common work dimensions. These scores represent the magnitude of the involvement of the dimension with each position.

When the organizational function is job evaluation, factor scores can be used to capture the pay policies of the organization. In this respect, the common factors function as compensable factors that are of value to the organization. When pay is regressed on the compensable factor scores, the resulting regression coefficients will yield an estimate of the impact each factor has in determining pay. Also, once the regression coefficients have been computed, they can be used to predict pay for each position.

Analysis of Dichotomous/Polychotomous Responses

The type of data analyzed with factor analytic techniques is important. Traditionally, many questionnaires have used a two-choice response format. If the item applies, it receives a score of one, and if it does not apply, it receives a score of zero. The Minnesota Multiphasic Personality Inventory (MMPI), The California Psychological Inventory (CPI), two of the five sections of the Strong-Campbell Interest Inventory, the Eysenck Personality Inventory (EPI), and the Rotter Internal-External Control of Reinforcement Scale (I-E Scale) use dichotomous scoring. According to Oswald and Velicer (1980), binary response sets were used because the scoring procedure for multiple response categories was too complicated and time consuming. Also, two-choice item formats (i.e., true-false, agree-disagree, yes-no) are frequently used because they require fewer and less complicated instructions, administration time is less, and they avoid scaling issues (Velicer & Stevenson, 1978). However, with the increasing availability of optical-mechanical scoring devices and computer scoring programs, the feasibility of performing graded multiple response category scoring has increased.

Several studies have evaluated solutions from factor and principal components analyses of data from two-choice and multiple-choice formats. Joe and Juliana (1973) examined the factor structure of the Rotter Internal-External Control of Reinforcement Scale (I-E) (Rotter, 1966) using dichotomous and polychotomous response categories. These researchers used two scoring systems. The respondents were asked to respond to a two-choice format and a graded multiple choice format which included six alternatives. The dichotomous and polychotomous responses were correlated and principal component analysis with a varimax rotation was used to analyze each correlation matrix. Results of the analyses revealed that although the same number of components were identified for the dichotomous versus polychotomous data, the six-point format produced a clear factor structure. The researchers asserted that the binary data did not permit the assessment of the degree of belief in internal or external control of reinforcement. They argued for the routine use of a 6-point response format in scoring the I-E scale.

Another examination of the component structure of a personality test was conducted by Velicer and Stevenson (1978). These researchers used principal component analysis with a varimax rotation to investigate the component structure of Form A of the Eysenck Personality Inventory. During a two week period, subjects completed two versions of the personality inventory. The standard version used a two-choice format and the modified version used a seven-choice response format. The analysis of the data from the traditional two-choice format revealed a two component solution. This solution accounted for 18% of the variance in the responses. However, 19 items of the inventory did not load on either component.

The analysis of the seven-point response format revealed a six component solution. This solution accounted for 46% of the observed variance. Also, item loadings were reported to be of larger magnitude with the seven-point format than with the two-choice format. The researchers concluded that the use of the seven-point format produced more meaningful relationships among items and resulted in better defined dimensions and more explained variance. Also, the researchers hypothesized that the reliability of the components should be better. This assertion is based on the assumption that the reliability of a component, as measured by coefficient alpha, is directly related to the size of its eigenvalue. Therefore, if components analysis of multiple-choice format data produces larger eigenvalues, then the reliability of these components should be better.

Oswald and Velicer (1980) investigated the component structure of Form B of the Eysenck Personality Inventory (EPI) with two-choice and seven-choice response formats. The seven-choice format was borrowed from the Comrey Personality Scales (CPS). A sample of 201 undergraduates responded to the EPI during two counter-balanced testing sessions which were one week apart. The resulting 57 X 57 intercorrelation matrix was used as input for principle components analysis with a varimax rotation. The results from the two-choice analysis revealed the traditional two component solution, neuroticism and extraversion. These components accounted for 16% of the variance. However, 23 items of the 57 item inventory did not load on either component and they did not appear to form an identifiable solution.

The results of the seven-choice analysis revealed a four component solution. The components were general anxiety, social extraversion, compulsivity, and affiliative concern. These components accounted for 29% of the variance. The researchers asserted that the seven-choice format allowed subjects to make more accurate and precise distinctions. The results of the Velicer and Stevenson (1978) and Oswald and Velicer (1980) studies appear to support the hypothesis that the seven-choice format allows for a more precise estimate of the underlying factors. The discrepancy in the number of components found in these two studies was attributed to using two different but not equivalent forms of the EPI.

Comrey and Montag (1982) investigated the component structure of the Comrey Personality Scales (CPS) using two-choice and seven-choice response formats. The CPS utilizes two seven-point scales. The anchors for the frequency scale are: "Never", "Very rarely", "Rarely", "Occasionally", "Frequently", "Very frequently", and "Always". The anchors for the probability scale are: "Definitely not", "Very probably not", "Possibly", "Probably", "Very probably", "Definitely". The researchers dichotomized the scales by setting responses 1 to 3 equal to 0 and responses 4 to 7 equal to 1. The analyses of the dichotomized and multiple response data using a varimax rotation revealed the same seven factor solution. However, the average intercorrelations among the items that defined each scale was greater for the seven-choice format than the two-choice format, .35 and .26 respectively. Also, the average factor loading for the seven-choice format was .52 while the average loading for the two-choice format was .44. The researchers concluded that the multiple-choice format allowed the subjects to more accurately express their self-descriptions. Apparently, the two-choice format forces the subject to make all or nothing discriminations that may not accurately represent the individual's behavior.

Response Format and Reliability

Many researchers have investigated the effect of the number of scale points on reliability. The results of the studies have generally been conflicting. Bendig (1954) and Matell and Jacoby (1971) concluded that reliability is not related to the number of scale categories. Jahoda, Deutsch, and Cook (1951) and Ferguson (1941) assert that reliability of a scale increases as the number of scale points increase. In an attempt to resolve this issue, Lissitz and Samuel (1975) performed a Monte Carlo study. Data were generated with three levels of covariance, .8, .5, and .2. The number of scale points used in the study were 2, 3, 5, 7, 9, and 14. Coefficient alpha, test-retest, and true, observed squared correlation reliability estimates were computed for the different sets of data. The results of this study demonstrated that formats that were constructed of more than two choices were consistently superior to the two-choice format. These researchers concluded that there is a increase in reliability when response formats include three or more response choices.

Item Response Theory

Another approach to scoring tests is item response theory (IRT). IRT is based on the assumption that the probability of an individual's response is directly related to an underlying characteristic of that individual (Hulin, Drasgow and Parsons, 1983). Because the underlying trait or latent trait can not be measured directly, IRT relates the probability of an individual's response to their standing on that trait. This relationship is described by an item characteristic curve (ICC). The level of analysis of IRT is the item. In order to generate accurate estimates of the level of the latent trait possessed by an individual, it is generally necessary to obtain the individual's responses to many items. Thus, IRT uses many observable behaviors to estimate the individual's standing on the latent trait which is usually referred to as theta, symbolized as Θ.

Assumptions of Item Response Theory

The use of IRT requires four assumptions: (a) unidimensional trait space; (b) local independence of item responses; (c) continuous distribution of latent trait; (d) the failure of an examinee to correctly answer items is solely a function of Θ and not caused by other factors such as insufficient time to complete the items. If a test is unidimensional, then the item response probabilities are a function of one latent trait. If the items are administered to different subpopulations, the distribution of item parameters should be the same. If they are not the same, then the test is measuring more than one latent trait (Hambleton & Swaminathan, 1985). However, according to Hulin et al., (1983) it is doubtful that any psychological instrument is exactly unidimensional. It is more realistic to view dimensions as composed of many subdimensions. Therefore, the question of unidimensional space concerns not whether the test is unidimensional or multidimensional but whether the test is sufficiently unidimensional to use IRT analysis. Lord and Novick (1968) asserted that the value of a model or assumption lies in its practical usefulness, and not in the ultimate truth regarding its dimensionality. Hulin et al., (1983) agreed with Lord and Novick but note that the dimensionality of a test may not be easily determined by a rational or logical inspection of its items. Some researchers argue that factor analysis should be used to investigate the dimensionality of a psychological instrument before IRT is used (Hulin et al., 1983, Hambleton & Swaminathan, 1985, and Lumsden, 1961, 1976). Lumsden recommends that the pool of test items, that have been selected on an empirical and rational basis, be analyzed using factor analytic techniques. Then the items that do not load heavily on the one dominant trait can be removed. This process is repeated until only those items that measure one factor are left.

Local independence refers to statistical independence between item responses from the same subject (Lord & Novick, 1968). In other words, how an examinee answers question one has no effect on how the examinee answers question two. The probability of a positive response to question two is independent of the probability of a positive response to question one. Local independence is based on the assumption that an examinee's response is solely a function of the individual's level of Θ. No other information concerning the examine is needed. All the information necessary to determine the probability of a positive response is contained in Θ and knowing the examinee's responses to additional items would not be helpful (Hulin et al., 1983).

Another assumption of IRT is that the latent trait is distributed continuously. According to Lord (1974) we do not have to assume that the latent trait is normally distributed. Therefore, the form of the continuously distributed trait does not have to be specified.

The last assumption concerns administration of the test under a timed condition. If the test is timed, the probability of a positive response will not only be a function of Θ but also of the speed at which the examinee is able to answer the questions (Hambleton & Swaminathan, 1985). Therefore, a test administered under a timed condition not only assess the examinee's level on the latent trait but how quickly they are able to respond to the items.

Item Response Theory Models

The exact form of the relationship between item responses and Θ is described by a model. The graphic representation of the relationship between the probability of an endorsed response and the level of Θ is illustrated by an item characteristic curve (ICC) (Tucker, 1946). According to Hulin et al., (1983), the best way to describe item parameters is graphically with an ICC.

The Guttman (1950) perfect scale model is a graphical representation that describes a perfect step function. An individual whose standing on Θ is less than a specific value will always fail to endorse the item. An individual whose standing on Θ is greater than a specific value will always endorse the item. Therefore the ICC, that graphically describes this relationship, takes the form of a step and the probability of a positive response is either 1 or 0. Some researchers assert that variables generally studied in the social science would rarely fit this model. The strict requirement that the probability of a positive response must be either 1 or 0 makes this model an unlikely representation of reality (Hulin et al., 1983).

Other models of the relationship between probability of a positive response to an item and the level of theta have been proposed. These models specify probabilities that range between 0 and 1. Lazerfield (1959) proposed a liner model of the relationship between the probability of a positive response and Θ. However, Torgerson (1958) notes that individuals who have low values of Θ can have negative probabilities and individuals with high scores on Θ can have probabilities greater than one.

The normal ogive has been suggested as a model for describing the relationship between the proportions of individuals who responded correctly to an item at various levels of Θ (Lawley, 1943).

$$P_i(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i(\theta-b_i)} e^{-y^2/2} d_y$$

The normal ogive model is used frequently in research on latent trait theory. Parameter a_i describes the item's ability to discriminate among different levels of theta. Graphically this can be seen as the magnitude of the slope of the ICC (see figure 2). Items that more sharply differentiate levels of Θ will have steep slopes. Parameter b_i is an index of the difficulty of an item and it is equal to the latent trait score at which half of the examines will answer the question correctly. Although the normal ogive model dominated much of the early research, it was subsequently replaced by logistic models, that provide good estimation of the normal ogive yet require much simpler computations (Crocker & Algina, 1986).

The one-parameter logistic model assumes that all items discriminate equally well. The probability of an examine responding correctly is therefore a function of item difficulty b_i and the examinee's level of Θ.

The two-parameter logistic model provides the best estimation of the normal ogive model. As with the normal ogive model, the two-parameter logistic model incorporates both item difficulty and item discrimination parameters. However, the fit of the normal ogive at its lower asymptote has not proven to be appropriate for data from multiple choice items. A better fit has been obtained by using an S-shape curve with a non-zero lower asymptote (Crocker & Algina, 1986). The lower asymptote for the one- and two-parameter models must approach zero at low levels of Θ. However, a lower asymptote considerably greater than zero is necessary to represent responses to multiple choice items, because even very low ability examinees can get difficult items correct (e.g., by guessing). Therefore, the three-parameter logistic model is used when guessing or dissimulation is possible. A "pseudo guessing parameter," c_i, is incorporated in the model to represent this nonzero lower asymptote of responding. As Lord (1974) noted, the value of c_i is frequently more than the probability of random guessing when examinees are able to eliminate obviously wrong multiple choice answers. When items have one or more multiple choice answers that are easily identified as incorrect by low ability examines, such items will discriminate less between low and high ability examines, and have higher c_i parameters.

Item Response Theory Versus Classical Test Theory

In order to better understand IRT it may be useful to compare it with classical test theory (CTT). IRT differs from classical test theory in several important ways. IRT item parameters (i.e., discrimination, difficulty, and "pseudo-guessing") are invariant across samples. Crocker and Algina (1986) refer to this property as "invariance to the selection of examinees". When the probability of a positive response is only a function of a single latent trait, then the ICC will have the same form in any subpopulation. Therefore, parameters can be estimated from one sample and with the use of a scaling constant can be applied to other samples.

Another important aspect of invariance to the selection of examines is person-free item calibration (Crocker & Algina, 1986). This means that item parameters can be estimated without having all the examinees answer all of the test items. This is accomplished by equating the responses of different groups. This transformation is used to place a_i and b_i on a common scale.

IRT defines the difficulty of an item as the level of Θ at which the examinees have a .5 probability of a positive response (Hulin et al., 1983). Therefore, an item's difficulty is defined relative to the level of the latent trait, not the characteristics of the sample which is used for item calibration.

According to CTT the difficulty of an item is defined as the proportion of correct responses p_i. Therefore, p_i is dependent on the sample. Crocker and Algina (1986) assert that for CTT an examinee's true score is contingent on the measuring process. An item's difficulty is a statistical concept whose value is the long run expected value. If an examinee were to repeatedly take the same test, without effects caused by learning or sensitization, the examinee's true score would be the average of the observed scores.

Measurement Error - Classical Test Theory

Although it is not possible to determine the precise amount of error in an specific examinee's observed score, CTT relies on assumptions regarding true scores and error scores to estimate the amount of error variance that can be expected in a group of examinee's observed scores. Assumptions of CTT concerning true and error scores are: (a) the mean of the error scores for a population will equal zero; (b) the correlation between true and error scores will be zero; (c) the correlation between error scores from two tests taken by the same examinees will be zero (Crocker & Algina, 1986). Therefore, since the true score is assumed to be a constant and the observed score is thought to be randomly distributed, then error scores must be randomly and normally distributed. An examinee's observed score is thus the sum of the true score and error score variances.

The relationship between true scores and observed scores is expressed in terms of the reliability index.

It is the ratio of the standard deviation of the true scores divided by the standard deviation of the observed scores.

$$\rho_{XT} = \frac{\sigma_T}{\sigma_X}$$

The square of this term is the reliability coefficient or test-retest reliability coefficient.

$$p_{x_1 x_2} = \frac{\sigma_T^2}{\sigma_X^2}$$

The standard error of measurement is an estimate the amount of error in an examinee's observed score. It is the average of the error standard deviations for a group of examines across repeated tests or parallel forms (Crocker & Algina, 1986). The test-retest reliability coefficient can be used to compute the standard error of measurement.

$$\sigma_E = \sigma_X \sqrt{1-\rho_{XX}}$$

Therefore, according to CTT a test has just one standard error of measurement or one estimate of reliability of a test; this value is constant regardless of the examinee's level of the trait.

Measurement Error - Item Response Theory

In IRT accuracy of estimates of Θ are based on the concept of information. Theoretically, the more information that exists about a parameter the more precise will be the estimate of that parameter. With regards to the ICC, the point at which the item provides the most formation is the point where the slope of the ICC is the greatest (i.e., where it most closely approximates a step function). This is also the point where the item works best in discriminating among different levels of Θ. Empirically, information is defined as the inverse of the squared standard error of Θ.

$$I_{\bar{x}} = \frac{1}{\sigma^2_{\bar{x}}}$$

An estimate of the accuracy of the computed Θs is given by the standard error of Θ. This is defined as one over the inverse of the information function. It is the standard deviation of the distribution of Θs for all individuals with a common Θ.

$$\sigma_{\theta} = \frac{1}{\sqrt{I(\theta)}}$$

Therefore, estimates of the standard error of Θ can vary at different levels of Θ.

In CTT the p_i is dependent on the characteristics of the sample. Therefore, there is no measure of item discrimination. IRT defines the discrimination of an item (two-parameter model) as the point where the ICC has its greatest slope. However, this does not apply precisely to the three-parameter model, because the value of c_i affects the discrimination of the item. The slope for a three-parameter model is dependent on the values of a_i and c_i. Therefore, there is an inverse relationship between the magnitude of c_i and the slope of the ICC. In other words, high c_i values reduce the information contained in the item responses.

Maximum Likelihood Estimation

Another important difference between CTT and IRT is that most IRT analyses use maximum likelihood estimation (MLE) to compute item parameters and Θ (Hulin et al., 1983). The basic rationale of MLE involves choosing values for item probabilities that maximize the likelihood function. There are three properties of MLE that make it an important method for the estimation of parameters. 1. Consistency of the parameters refers to the agreement between the value of the estimators and value of the population parameters. The estimated parameters will be consistent with the population parameters when estimates are computed from large samples. 2. Asymptotic normality specifies that the sampling distribution will reach normality as the sample size increases. 3. Asymptotic efficiency stipulates that sampling variance will be minimal for large samples. According to Hulin et al., (1983) these properties mean that the estimates will converge to the population parameters as the sample size increases.

In testing situations neither the examinee levels of Θ nor the item parameter are known. Lord (1980) offered an iterative procedure for simultaneously estimating Θs and item parameters. Basically this method involves assigning initial estimates to the parameters and letting the likelihood equations adjust these values until they converge. First, item parameters are estimated while holding latent trait estimates constant and then item parameters are held constant while Θs are adjusted. This iterative process continues until the estimated values of Θ and the item parameters change by a small arbitrarily set value. According to Hulin et al., (1983), the simultaneous iteration process appears to yield estimates of parameters that are accurate enough for most applications.

Item Response Theory Versus Factor Analysis

Some researchers have compared factor analysis with IRT. Takane and de Leeuw (1987) offered a formal proof for the equivalence of MLE of the two-parameter normal ogive model in IRT and factor analysis for dichotomized data. According to these authors, for the special case of dichotomous data, IRT and factor analysis are equivalent formulations of the same model (p. 395). They assert that the only difference between these approaches is where the marginalization is performed.

Takane and de Leeuw also note that the relationship they demonstrated between IRT and factor analysis should also hold for the analysis of graded polychotomous data. They assert that the MLE of the normal ogive model for graded scores (Samejima, 1969) is equivalent to factor analysis of ordered categorical data which was proposed by Muthen (1984).

Harvey (1989) empirically investigated latent trait versus factor scoring of the Job Element Inventory (JEI) (Cornelius & Hakel, 1978). The JEI is a 153-item worker-oriented job analysis questionnaire which is completed by job incumbents or supervisors. Workers respond to the items by marking a six-choice, relative-time-spent scale.

IRT analysis using the three-parameter logistic model was compared to common factor analysis of dichotomized data. This study had two major objectives: first, to determine the form and degree of the relationship between factor scores and latent trait scores and, second, to investigate whether factor analytic and IRT methods will identify the same items as important to the assessment of common work dimensions.

In order to use the three-parameter logistic model, graded ordered relative-time-spent data from 1,767 positions were dichotomized. The response format of the JEI is a six-choice graded frequency scale: responses 0 to 1 were assigned a value of 0; responses 2 through 5 were assigned a value of 1. A common factor analysis of the dichotomized data was performed with an orthoblique rotation. This type of rotation allows the common factors to become correlated if needed to fit the sample data. The three-parameter logistic IRT model was specified using the Logist computer program (Wingersky, Barton, and Lord, 1982).

The results of the common factor analysis of the dichotomized data revealed 6- or 16-factor solutions. Items from the first 4 factors of the 6-factor solution were the subject of IRT analyses.

The IRT analyses revealed several findings. First, the discrimination parameters of the ICCs for the JEI items had generally high values. This indicates the these items are capable of discriminating well between different levels of the dimension. However, an inspection of the test information functions and standard errors of Θ for each dimension revealed that the dimensions worked best only for jobs with an average to moderately-above-average level of the dimension. Positions that have relatively high or low values of Θ on the dimensions will be measured with much less precision when using dichotomous data in the three-parameter model. As Harvey (1989) pointed out, for a 95% confidence interval, (i.e., 1.96 X the standard error of Θ) which is a relatively large degree of error, dimension 1 accurately measures jobs from -.80 to 2.5, dimension 2 from -1.2 to 2.0, dimension 3 from .5 to 2.2 , and dimension 4 from -1.4 to 1.8. Therefore, positions that score outside these ranges will be measured with unacceptably high levels of error.

Second, for many items the relationship between the probability of a positive response to an item and the level of the dimension is nonlinear. In many cases, the a_i parameter caused the ICC to approximate a step function.

Third, the value of the c_i parameter was frequently quite high. For some items, 25% to 50% of the raters endorsed the item regardless of having very low levels of the work dimension. When the value of the c_i parameter is high, the items ability to discriminate is greatly reduced.

Fourth, the degree and form of the relationship between factor scores and latent trait scores is strong and nonlinear. Harvey (1989) found that the correlations between the factor scores and latent trait scores were all significant at the .01 level and ranged from .79 to .91. However, at high and low values of factor scores the form of the relationship between factor scores and latent trait scores was distinctly nonlinear. Harvey interpreted this to mean that Θ scores may discriminate better than factor scores at high or low values (e.g., with respect to predicting external variables of interest).

Overall, the results of this investigation demonstrated that factor analytic and latent trait methods of scoring a worker-oriented job analysis instrument yield different results. If this information were used for job evaluation, it is plausible that these approaches would yield different predictions of pay.

Number Endorsed Scoring

Another approach to scoring worker-oriented job analysis instruments is a simple number endorsed method. Dimension scores could be determined from a simple count of the number of positively endorsed items for each dimension. This would yield a dimension score for each examinee and give an estimate of the amount of the dimension that is involved in each position. As with factor scores or latent trait scores, these dimension scores could be used to estimate the value of the position to the organization. In its simplest application, number endorsed sores could be given unit weights (all predictors are weighted by 1.0) in a multiple regression equation to predict pay. Cascio, Valenzi, and Silbey (1978), Lawshe and Schucker (1959), Wainer (1978) assert that unit weighting can perform as well or better than optimal weighting. Einhorn and Hogarth

(1975) argue that unit weighting has three advantages over optimal weighting: (a) Since unit weights are not estimated from the data, they do not reduce the degrees of freedom and decrease power of the statistical test; (b) no error is involved in the estimation of the weights; (c) weighting can not cause error in the estimates by reversing the true weights of the predictors.

It is conceivable that unit weighting could prove superior to weights derived from policy capturing. The utility of the dimension scores in predicting pay is based on their correlation with the criterion and not in how well they explain the intercorrelations of the items. If policy capturing weights are used, it is conceivable that this could reduce the predictive value of an item even though the item is highly correlated with the pay. This would occur when the predictor is given relatively small weights.

Present Study

Objectives

The present study investigate the usefulness of three scoring approaches, two response formats, and two response scales for job evaluation. It seems plausible that these factors may affect the accuracy of estimates of pay made from a worker-oriented job analysis questionnaire, the CMQ. Specifically, the objectives are to: (a) compare the usefulness of factor scores, latent trait scores, and number endorsed scores computed from polychotomous data and dichotomous data; (b) test whether ordered frequency scale ratings are more useful than relative importance scales ratings in estimating factor scores; (c) determine the nature of the relationship between factor scores and latent trait scores; (d) compare the precision of latent trait scores computed from polychotomous data and dichotomous data; (e) test whether the reliability of factor scores is affected by response format.

Hypotheses

Based on the objectives of this study, eight hypotheses will be tested:

Hypothesis 1, Factor Scores and Polychotomous versus Dichotomous Data. Factor scores computed from order polychotomous frequency data will prove to be better predictors of pay than factors scores computed from dichotomous data. Multiple choice formats give the examinee more opportunity to respond accurately to the items of the job analysis instrument. Dichotomous response formats force the examinee to report extreme levels of the behavior. This may not accurately describe the job related behaviors. Therefore, there is more accurate information in the polychotomous data. Since the predictive value of the factor scores is dependent on the quality of the data used to construct the correlation matrix, the utility of the factor scores will be greater when polychotomous data is used.

Hypothesis 2, Latent Trait Scores and Polychotomous Versus Dichotomous Data. Latent trait scores computed from polychotomous frequency data will prove to be better predictors of pay than latent trait scores computed from dichotomous data. Estimates of Θ and item parameters are dependent of the quality of the item responses. Therefore, data which better reflect the true nature and extent of individual differences will have greater predictive value.

Hypothesis 3, Factor Scores and Relative Importance Scale Versus Multiple Category Frequency Scale Data. Factor scores computed from polychotomous frequency data will be superior to factor scores computed from polychotomous relative importance data as predictors of pay. Workers can report more accurate estimates of behavioral frequency (eg., hourly to many times an hour, daily) than relative importance (eg., very minor importance, below average importance). Therefore, the worker's response on the frequency scale will better represent the true extent of individual differences.

Hypothesis 4, Latent Trait Scores and Relative Importance Versus Multiple Category Frequency. Latent trait scores computed from polychotomous frequency data will be a superior to latent trait scores computed from polychotomous importance data as a predictor of pay. Frequency data will be more representative of what the examine does on the job and therefore provide superior estimates of Θ.

Hypothesis 5, Number Endorsed Scoring Versus Factor Scoring. Dimension scores computed by number endorsed scoring will be superior, in terms of prediction and estimation of pay, to factor scores or latent trait scores in predicting and estimating pay. The predictive value of each item is based on its correlation with the criterion. Weighting items with factor pattern weights could diminish or reverse the predictive value of an item. This would occur when the item is correlated with the criterion but is given a small weight or a reversed sign in the multiple regression equation.

Hypothesis 6, Relationship of Factor Scores to Latent Trait Scores. When factor scores are plotted against latent trait scores, the resulting distribution will be nonlinear for dichotomous and polychotomous data. Although some researchers (eg., Takane and de Leeuw) assert that IRT and factor analysis are equivalent techniques, Harvey (1989) and Parker (1991) found that the relationship between factor scores and latent trait scores was nonlinear for dichotomous response to the JEI.

Hypothesis 7, Precision of Latent Trait Scores Estimated From Polychotomous Versus Dichotomous Data. Standard errors of Θ will be lower over a greater range of Θ when latent trait scores are computed form polychotomous data than dichotomous data. Since multiple choice formats provide the examinee with a more accurate opportunity to report job related behavior, there will be less error in these estimates.

Hypothesis 8, Reliability of Factor Scores Estimated from Polychotomous Versus Dichotomous Data. Reliability of factor scores will be greater when factor scores are computed from polychotomous data versus factor scores computed from dichotomous data. The reliability of factor scores will be of larger magnitude when computed from more accurate estimates of individual differences.

METHOD

Data

The CMQ standardization data base contains the responses of 2796 job incumbents. Organizations in the data base include an insurance company, hospital, and a large utility company. The data consist of five-choice graded responses from several response scales such as "Frequency", "Importance", "Who Initiates Contacts" scales. The five-choice frequency and importance scales were dichotomized by assigning a value of 0 to responses coded 0 and a value of 1 to response coded 1-5.

Instrument

The Common Metric Questionnaire (CMQ), version 1.12, is a worker-oriented job analysis questionnaire which is used for several human resource management functions such as, job classification, job evaluation, and performance appraisal. The CMQ contains more than 2000 items which are divided into rationally derived categories (e.g., working in demanding personal situations, making decisions, and managing operations and production). The CMQ takes approximately 1 to 2 hours to complete.

Procedure

Organizations were offered an opportunity to participate in the standardization of the CMQ. Each organization was instructed to administer the instrument to their job incumbents and to have the worker's supervisor or a designated personnel officer check each CMQ booklet. This procedure was necessary in order to insure that the job incumbent's responses and reported pay were accurate. However, one organization in the data base, which contributed approximately 71% of the observations, administered the CMQ anonymously. Therefore, the job incumbents' responses and reported pay could not be verified by the supervisors or the researchers.

Instructions in the CMQ booklet ask the worker to answer all questions which apply to his/her current job, not to worry that many of the questions may not apply to their job, and to give a current description of the job. All organizations received a report which described the results of the job analysis and job evaluation.

Factor Analysis

In order assess the dimensionality of the item pools furnished by Harvey (in press) an exploratory factor analysis of each of the six item pools (e.g., manual labor, middle management, external contacts, upper-level management, using math and figures, supervisory duties) was completed. Separate analyses were completed for the polychotomous and dichotomous responses. A common factor model with maximum likelihood estimation and Harris-Kaiser p=.5 orthoblique rotation was used.

After it was determined that the item pools were sufficiently unidimensional, a principal components model with a orthogonol rotation was used to compute the factor scores. Factor scores were estimated using the SAS (SAS Institute, Inc., 1987) version 6.04 PROC SCORE. This procedure computes factor scores by multiplying each examinee's raw item scores by each item's scoring coefficient and summing these products to form a linear composite.

IRT Analysis

The polychotomous data were analyzed using the graded response model (Samejima, 1969). Latent trait scores and parameter estimates were completed using the Multilog computer program (Thissen, 1988). Separate estimates of latent traits were computed from the dichotomous and polychotomous responses of the frequency and importance scales.

The graded response model (Samejima, 1969) uses the entire response pattern in estimating latent trait scores. An item characteristic curve (ICC) is computed for each endorsed response and the computation considers each response in relation to the other categories. If P*(k) is the response in category k or above, then the probability of a response in category k is

$$P^*(k) - P^*(k+1)$$

Responses at the extremes of the scale provide relatively more information (i.e., never, often) than responses in the middle of the scale (i.e., sometimes). Although this model was developed in 1969, the availability of high speed computers and numerical algorithms has only recently made the computation of its complicated calculations feasible.

A logistic two-parameter model was used for the dichotomous data. This is functionally equivalent to the graded response model for dichotomous data (Thissen, 1988). Graded responses were dichotomized so that item parameters and latent trait estimates could be computed using the two-parameter logistic model.

Number Endorsed Scoring

Simple summated scores were computed for each dimension. These scores were computed by summing the raw item scores. Values of 0-5 were used for the polychotomous data and values of 0-1 were used for the dichotomous data.

Pay

An examination of the data revealed 112 observations with extreme or missing values of pay. Reported pay values ranged from $0.00 to $20,000,000.00. Since these extreme values are unrealistic and the values of pay for these observation could not be verified, only observations with reported pay in the range from $8,000.00 to $100,000.00

were used for the analyses. It was thought that this range would include most of the accurate reports of pay. Accordingly, 33 observations with pay less than $8000.00, 56 observations with pay greater than $100,000.00, and 33 observations with missing values for pay were excluded from the analyses.

RESULTS

Hypothesis One

Six sets of items (scales) were provided by Harvey (personal communication, March 16, 1991) for the analyses. The results of the factor analyses of the polychotomous and dichotomous data from the frequency scale revealed a one-factor solution for each of the six scales. The proportion of variance accounted for by the first factor of each of the seven scales suggests that each dimension is relatively unidimensional (see Table 1).

In order to assess the relationship among factor scores and pay, factor scores from polychotomous data were correlated with pay. Although all of the correlations between pay and the factor scores were significant, only the factor scores from four scales had correlations of sufficient magnitude to demonstrated <u>meaningful</u> relationships with pay. The correlations for scale one and three with pay were -.07 and .09 respectively. Therefore only scales two, four, five, and six were used for the analyses (see Table 2). Descriptive statistics of the factor scores for these scales are shown in Table 3. Table 4 contains a list of the CMQ items for each scale.

In order to evaluate the predictive value of factor scores computed from polychotomous and dichotomous data, regressions of pay on factor scores were performed. Factor scores and pay were plotted for each scale and type of data. However, a visual examination of the scatterplots failed to reveal the form of their relationship (see Figures 1-8). Therefore nonlinear prediction terms were added to the multiple regression equations. Multiple regressions were performed using the SAS (1987) proc regression selection equals stepwise procedure.

The results of the linear and nonlinear regressions of pay on factor scores computed from polychotomous data revealed that scale 2, 4, 5, and the quadratic term for scale 2 reliability accounted for .2352 of the variance in pay (see Table 5). Although the stepwise regression procedure continued to add quadratic and cubic terms into the equation, the increase in multiple R^2 was less than .01. It was decided to stop entering additional predictors into the regression equations when the resulting change in multiple R^2 was less than .01.

The results of the multiple regression of pay on factor scores computed from dichotomous data demonstrated that scale two, four, and five accounted for .2658 of the variance in pay, (see Table 5). Again, predictors which contributed less than .01 to the increase in multiple R^2 were not entered into the regression equation.

In order to evaluate the utility of factor scores computed from polychotomous versus dichotomous data, the residuals (actual pay - estimated pay) for each type of data were computed. The mean absolute residuals for factor scores computed from polychotomous and dichotomous data were $8676.12 and $8531.02 respectively (see Table 6).

In order to assess whether factor scores computed from polychotomous data were superior to factor scores computed from dichotomous data for estimating pay, the difference in the mean absolute residuals was tested to determine whether it was significantly different from zero. The results of the paired *t*-test demonstrated that estimates of pay by factor scores computed from dichotomous data are superior to factor scores computed from polychotomous data (see Table 7).

Hypothesis Two

In order to evaluate the predictive value of latent trait dimension scores computed from dichotomous versus polychotomous data, latent trait scores for each of the four scales and type of data were computed (see Table 8). Latent trait scores and pay were plotted for each scale and type of data (see Figure 9-16). A visual examination of the scatterplots failed to reveal the form of this relationship. Therefore, nonlinear prediction terms were added to the regression equations and pay was regressed on the scale scores. The multiple regressions were conducted according to the same procedure used for Hypothesis One.

The results of the regression of pay of latent trait scores computed from polychotomous data revealed that latent trait scores for scales two, four, and five accounted for .2522 of the variance in pay (see Table 9).

The regression of pay on latent trait scores computed from dichotomous data revealed that latent trait scores computed from scales two, four, and five accounted for .2666 of the variance in pay, (see Table 9).

The utility of latent trait scores computed from polychotomous and dichotomous data was evaluated by computing the mean absolute residual (actual pay - estimated pay) from the regression analysis. The mean absolute residuals were $8610.71 and $8548.01 respectively (see Table 10).

In order to estimate the value of latent trait scores computed from polychotomous data versus dichotomous data in estimating pay, a paired t-test was used to test whether the difference in the mean absolute residuals is significantly different from zero. The results of this analysis demonstrated that latent trait scores computed from dichotomous data are superior to latent trait scores computed from polychotomous data for estimating pay (see Table 11).

Hypothesis Three

Many of the items of the scales provided by Harvey (personal communication, March 16, 1991) did not have importance scale ratings. However, the data base did contain a pool of items in the category "Using Information" which did have frequency and importance scale ratings (scale seven). This rationally derived set of items was used to test hypotheses three and four (see Table 12). Factor scores were computed for each type of data (see Table 13). The correlations of factor scores and pay for polychotomous and dichotomous data were .25 and .33 respectively.

In order to evaluate the form of the relationship between pay and factor scores estimated from the frequency and importance scales, scatterplots were constructed (see Figures 17-20). However, a visual inspection of the scatterplots failed to reveal the form of their relationship. Therefore, nonlinear regression terms were included to the regression equations. The multiple regressions were conducted according to the same procedure used for Hypothesis One.

The regression of pay on factor scores computed from the frequency responses and polychotomous data demonstrate that scale seven and its cubic term accounted for .0839 of the variance in pay (see Table 14). The regression of pay of factor scores estimated from the importance scale responses and polychotomous data revealed that scale seven and its cubic term accounted for .0891 of the variance in pay (see Table 14).

The results of the regression of pay on factor scores computed from the frequency scale and dichotomous data revealed that scale seven and its quartic term reliability accounted for .1321 of the variance in pay (see Table 14). Factor scores estimated from the importance scale and dichotomous data reliability accounted for .1331 of the variance in pay (see Table 14).

In order to evaluate the utility of factor scores computed from ordered multiple frequency scale data versus ordered relative importance scale data in estimating pay, the mean absolute residuals (actual pay - estimated pay) for each scale and type of data were computed (see Table 15). The mean absolute residuals from the frequency and importance scales, polychotomous data, were \$8676.12 and \$8610.71 respectively. The mean absolute residuals from the frequency and importance scales, dichotomous data, were \$8531.02 and \$8548.01 respectively.

In order to estimate the value of factor scores estimated from frequency and importance scales in estimating pay, a paired _t_-test was used to test whether the difference in the mean absolute residuals was significantly different from zero (see Table 16). The results demonstrated that factor scores estimated from the importance scale are superior to scores computed from the frequency scale for estimating pay for either type of data.

Hypothesis Four

In order to evaluate the predictive value of latent trait dimensions scores computed from ordered multiple frequency scale versus ordered relative importance scale data, latent trait scores were estimated for each scale and type of data (see Table 17). The correlations for these variables are shown in Table 18.

In order to estimate the form of the relationship between pay and factor scores computed from each scale and type of data, scatterplots were constructed (see Figures 21-24). A visual examination of the scatterplots failed to reveal the form of their relationship. Therefore, nonlinear regression terms for each scale were included in the regression equations. The analyses were conducted according to the same procedures as for Hypothesis One.

The regression of pay on latent trait scores estimated from the frequency scale and polychotomous data demonstrated that scale seven and its quadratic and cubic terms reliability accounted for .0872 of the variance in pay (see Table 19). The regression of pay on latent trait scores estimated from the importance scale and polychotomous data revealed that scale seven and its quadratic and quartic terms reliability accounted for .1240 of the variance in pay (see Table 19).

The regression of pay on latent trait scores estimated from the frequency scale and dichotomous data revealed that scale seven and its cubic term accounted for .1281 of the variance in pay (see Table 19). The regression of pay on latent trait scores computed from the importance scale and dichotomous data demonstrated that scale seven and its cubic term accounted for .1328 of the variance in pay (see Table 19).

In order to evaluate the utility of latent trait dimension scores computed from ordered multiple frequency scale data versus ordered relative importance scale data in predicting pay, latent trait scores from each scale and type of data were used to estimate pay (see Table 20). The mean absolute residuals estimated from the frequency and importance scales and polychotomous data were $9237.66 and $9012.25 respectively. The means absolute residuals estimated from the frequency and importance scales and dichotomous data were $9047.38 and $8998.79 respectively.

In order to estimate the value of latent trait scores estimated from frequency and importance scales in estimating pay, a paired _t_-test was used to test whether the difference in the mean absolute residuals was significantly different from zero (see Table 21). This analysis revealed that latent trait scores computed from the importance scale and polychotomous or dichotomous data are superior to latent trait scores computed from the frequency scale for estimating pay.

Hypothesis Five

In order to evaluate the predictive value of number endorsed scale scores of polychotomous and dichotomous data and to test whether number endorsed scoring would be superior to factor scoring, number endorsed scores for each dimension and type of data were computed (see Table 22). Correlations for these variables and pay are shown in Table 23.

Number endorsed scores and pay were plotted and the scatterplots were visually examined in order to determine the form of their relationship (see Figures 25-32). However, this examination failed to reveal the form of their relationship. Therefore, nonlinear terms were included in the regression equations and the analysis was performed according to the same procedure for Hypothesis One.

The results of the stepwise regression of pay on number endorsed scores computed from polychotomous data revealed that number endorsed scores from scale two, four and the quadratic term for scale two, accounted for .2716 of the variance in pay (see Table 24).

The results of the regression of pay on number endorsed scores computed from dichotomous data demonstrated that number endorsed scores for scale two, scale four, and scale five accounted for .2716 of the variance in pay, (see Table 24).

The utility of number endorsed scores computed from polychotomous and dichotomous data was assessed by computing the absolute residual (actual pay - estimated pay). The mean absolute residuals for the polychotomous and dichotomous data were $8840.84 and $8501.64 respectively (see Table 25).

In order to estimate the value of number endorsed scores estimated from polychotomous versus dichotomous data in estimating pay, a paired *t*-test was used to test whether the difference in the mean absolute residuals was significantly different from zero (see Table 26). The results demonstrated that number endorsed scores computed from dichotomous data are superior to number endorsed scores computed from polychotomous data in estimating pay.

The utility of number endorsed scoring versus factor scoring of polychotomous and dichotomous data was evaluated by using a paired *t*-test to test whether the difference in the mean absolute residuals was significantly different and the quadratic term for scale two, accounted for .2716 of the variance in pay (see Table 24).

The results of the regression of pay on number endorsed scores computed from dichotomous data demonstrated that number endorsed scores for scale two, scale four, and scale five accounted for .2716 of the variance in pay, (see Table 24).

The utility of number endorsed scores computed from polychotomous and dichotomous data was assessed by computing the absolute residual (actual pay - estimated pay). The mean absolute residuals for the polychotomous and dichotomous data were \$8840.84 and \$8501.64 respectively (see Table 25).

In order to estimate the value of number endorsed scores estimated from polychotomous versus dichotomous data in estimating pay, a paired $\underline{t}$-test was used to test whether the difference in the mean absolute residuals was significantly different from zero (see Table 26). The results demonstrated that number endorsed scores computed from dichotomous data are superior to number endorsed scores computed from polychotomous data in estimating pay.

The utility of number endorsed scoring versus factor scoring of polychotomous and dichotomous data was evaluated by using a paired $\underline{t}$-test to test whether the difference in the mean absolute residuals was significantly different form zero (see Table 26). The results were mixed. Factor scores were superior to number endorsed scores when computed from polychotomous data. Number endorsed scores were superior to factor scores when computed from dichotomous data.

Number endorsed scoring was also compared to latent trait scoring for estimating pay (see Table 26). A test of the mean absolute residuals revealed that latent trait scores were superior to number endorsed scores when computed from polychotomous data. Number endorsed scores were better predictors of pay than latent trait scores computed from dichotomous data.

Hypothesis Six

In order to determine the form and degree of the relationship between factor scores and latent trait scores, factor scores and latent trait scores for each scale and type of data were correlated (see Table 27). Factor scores and latent trait scores were also plotted (see Figures 33-40).

A visual examination of the scatterplots revealed a curvilinear relationship for most of the scales. This relationship is not limited to extreme scores but appears to apply to scores nearer the means. In order to assess the form and degree of this relationship, linear or nonlinear regression of factor scores on latent trait

Table 30. Scatterplots of factor and three-parameter latent trait scores were constructed (see Figures 41-44).

Regressions of factor scores on latent trait scores demonstrated that the relationship between factor scores and latent trait scores is cubic for scale two, (see Table 31). However, the relationship between factor scores and latent trait scores for scales four is quadratic and linear for scales five and six (see Table 31). The results of the present study support the finding of Harvey (1989) and Parker (1991) that there is a curvelinear relationship between factor scores and latent trait scores estimated with the three-parameter model.

When latent trait scores computed from the three-parameter logistic model were used to estimate pay, scales two and five accounted for .2059 of the variance in pay (see Table 32). The mean absolute residual, actual pay - predicted pay, was $15453.40 (see Table 33).

In order to investigate whether latent trait scores computed from a three-parameter logistic model would be superior to latent trait scores computed from a two-parameter model, the difference in the mean absolute residuals was tested to determine if it was significantly different from zero (see Table 34). The results reveal that latent trait computed from the two-parameter logistic model are superior to latent trait scores computed from the three-parameter model.

Hypothesis Seven

In order to evaluate the precision of latent trait scale scores computed from polychotomous versus dichotomous data, test information functions and standard errors of Θ for each scale were computed and plotted against Θ (see figures 45-52). Estimates of the scale parameters, from the two-parameter and graded response models are shown in Table 35. The height and width of the test information functions illustrates the amount of information that the scales provide at different levels of Θ. The plots of the standard error of theta reveals the accuracy of estimates of Θ over a range of Θ.

A examination of the test information functions for each scale estimated from polychotomous and dichotomous data illustrates that there is substantial overlap in the area described by the curves. However, there is considerable differences in the shape of the curves (e.g., scale four, five, and six). The figures demonstrate that latent trait scores estimated from polychotomous versus dichotomous data provide different amounts of information at common levels of Θ.

An examination of the standard error of Θ for each scale reveals the precision of the estimates of Θ. In order to quantify and compare the level of precision of latent trait scores estimated from polychotomous versus dichotomous data, a 95% confidence interval of plus or minus 0.5 trait units was established. This interval represents a 95% confidence level that an individual's true latent trait score lies in the interval around the estimated Θ.

The results reveal considerable overlap in the range of accurate measurement for latent trait scores for scale two, four, and six. (see Table 36). However, no values of Θ for scale five are measured with this degree of accuracy. A comparison of the standard errors for latent trait scores estimated from polychotomous versus dichotomous data revealed that the intervals of accurate measurement (95 percent confidence interval) were 0 to 2.2 _z_ units, _M_=1.17 for the polychotomous data and 0 to 2.2 _z_ units, _M_=1.25 for the dichotomous data. This demonstrates that the four scales measure, with an acceptable level of accuracy, a relatively small range of Θ. If the ranges of accurate measurement are summed for the four CMQ scales, latent trait scores estimated from dichotomous data provide accurate estimates of Θ over a greater range of Θ than estimates from polychotomous data, 5.0 and 4.7 _z_ units, respectively.

Harvey (1989) computed the standard error of theta for four scales of the Job Element Inventory (JEI) and found that the confidence intervals (95 percent) ranged from 2.6 to 3.3 _z_ units, _M_ =3.1. This researcher stated that this demonstrates that the JEI items measure a narrow range of Θ.

The narrow band of accurate measurement found for the CMQ in the present study is due to the relatively high discrimination parameter (_a_) for each scale. The mean of the _a_s for the CMQ scales ranged from 1.52 to 2.49 _z_ units. Hulin et al., (1983, pp. 80-81) found that _a_s for items of the Job Descriptive Index (JDI) ranged from 0.20 to 1.21 _z_ units. Harvey (1989) reported that mean of the _a_s for the four scales of the JEI ranged from 0.93 to 1.58 _z_ units. Therefore, the large discrimination values for the CMQ items cause the test information functions to be sharply peaked thus providing a high degree of precision over a narrow range of Θ.

Hypothesis Eight

In order to evaluate the reliability of factor scores computed from dichotomous versus polychotomous data, reliability estimates were made using a measure of internal consistency of the factor scoring coefficients, derived by Kaiser and Michael (1977):

$$KR\ 20(f) = \left(\frac{p}{p-1}\right)\left(1-\sum_{j=1}^{p} W_{jf}^{2}\right)$$

where f is the number of factors, p is the number of variables in the linear combination, and W_{fj}^{2} is the weights (factor scoring coefficients).

The estimates of reliability of each scale reveals that factor scores computed from dichotomous data are more reliable than factor scores computed from polychotomous data for scales two, five, six and seven (see Table 37). However, reliability estimates for scale four demonstrated that factor scores computed from polychotomous data are more reliable than factor scores estimated from polychotomous data.

In order to assess the reliability of number endorsed scores estimated from the frequency and scale and polychotomous and dichotomous data, coefficient alpha was computed for each scale (see Table 37). The results reveal that number endorsed scores computed from dichotomous data were more reliable than number endorsed scores estimated from polychotomous data for scales four, five, and six. However, number endorsed scores for scale four were more reliable when estimated from polychotomous data versus dichotomous data.

DISCUSSION

This study demonstrated that scale scores computed using dichotomous data were superior predictors of pay. These results do not support the hypothesis that polychotomous data contain more information about the positions such that scaled scores from polychotomous data will be superior predictors of pay. However, this hypothesis assumes that the additional information, variance in the scale scores, is accurate. If the additional information on which the scales scores are based were inconsistent with the workers' job related behaviors, the increase in variance of the scale scores would reflect an increase in error variance. If the increase in variance of the predictors occurs due to an increase in error variance, the predictive value of the scale scores, estimated by any method, would be diminished.

Unfortunately, the information contained in the CMQ data base does not allow an investigation of the accuracy of the additional information. There also have been no reported data on the accuracy of graded response frequency and importance scale ratings by job incumbents to a worker-oriented job analysis questionnaires. The CMQ is the first worker-oriented instrument to rely solely on ratings by job incumbent. The Position Analysis Questionnaire is completed by job analysts who interview or observe multiple subject matter experts (e.g. supervisors, managers, or job incumbents) and make judgements about each position.

Several researchers have questioned the accuracy of job incumbent ratings and have reported high levels of within job variance. Green and Stutzman (1986) found that job incumbents are not equally accurate in their responses to a task-oriented job analysis inventory. Some workers rated many tasks as important even when they were not related to the job. Green and Stutzman assert that job analysts should not assume that errors in the data will be random and will cancel out when large samples are used. Instead, it appears that job incumbents often distort frequency and importance scale ratings in order to emphasize or deemphasize certain aspects of their positions.

Harvey and Wilson (1990) demonstrated that dichotomous, "Do-You-Perform" task judgements, contain as much information about the job as polychotomous relative-time-spent frequency ratings do. These researchers reported that job incumbents within the same job title often disagreed on objective job tasks and that this disagreement created high within job variance.

Other researchers have identified several reasons why workers within the same job title give different descriptions of their positions. Cascio (1987) points out that questionnaires may contain items that are ambiguous or interpreted differently. Differences among workers in their level of cooperation or motivation could also increase within job variance.

Harvey, (1989) asserts that certain workers may intentionally emphasize certain aspects of their positions in order to present a better picture than exists. This researcher suggests including "lie" scales or employing more sophisticated IRT appropriateness measurement techniques to identify those workers who have insufficient motivation, verbal skills, or who consciously distort their ratings.

Christal (1974) asserts that many workers are unable to give precise estimates of the amount of time they engage in certain work activities. Perceptual differences among workers may affect their frequency and importance scale ratings.

If workers are unwilling or unable to give accurate information about their positions, it seem plausible that the response format could have an effect on the workers' ratings. A dichotomous format discourages blatant attempts to distort the ratings and simplifies the choices. If workers are forced to chose between "I perform the behavior" or "I do not perform the behavior", that choice may decrease within job variance. Conversely, use of a polychotomous format may permit workers to make subtle distortions of their positions and the multiple alternative format may cause some confusion.

Even when frequency and importance scale ratings are checked by the worker's supervisor, the polychotomous format leaves room for interpretation. For example, if two assembly line workers with the same job title and with no consulting, supervisory, or managerial duties are asked "In order to perform your job, do you attend meetings to consult or give specialized information?", they may give different answers. One worker may consider the monthly department meeting as an opportunity to give specialized information about their position and respond by answering yes and marking "Monthly". The other worker may not consider these meetings "consulting" and answer no. These workers would receive scores of 3 and 0, respectively. If the two workers were asked the same question with a dichotomous format, they would both probably answer no and receive a score of zero. The dichotomous format may inhibit the first worker, who answered yes with the graded response format, from stretching the point. If a worker is motivated to distort their responses or is confused by the multiple alternatives, use of the dichotomous format may delimit these distortions while using a polychotomous format may encourage them.

There appear to be many possible reasons why workers with the same job title rate their job differently. Although Christal, 1974; Cascio, 1987; Harvey, 1989 believe that these differences are a product of variations in perceptual functioning, verbal skills, levels of motivation, or a conscious desire to exaggerate their jobs, there may be other reasons why workers rate their positions differently. Workers with in the same job title may be performing the job in different ways or they may have different job duties.

Cascio (1987) reports that person-determined changes are frequent and occur when workers in the same job accomplish their job related objectives in different ways. If this occurred, worker- or task-oriented job analyses would group these workers into different jobs classifications.

Stutzman (1983) found statistically significant differences among positions in a task-oriented job analysis of a single job classification. This researcher noted that these differences would have a meaningful effect on the results of selection, training, performance appraisal, and job classification programs. However, the design of this study did not allow the researchers to determine whether the differences in the ratings reflected actual differences among the positions.

One of the purposes of a job analysis, job classification, is to assign a common job title to workers who perform the same job activities. It is widely known that job analysts frequently find that workers are misclassified and that their job titles need to be revised. Therefore, workers within the same job title may be engaged in different work activities.

It also seems plausible that workers who have the same job title but who are performing their positions in different ways, may be receiving the same pay. A worker's pay may be more a function of the worker's job title than their present work activities. The present study used workers' self reports of their work related behavior in order to predict pay. However, most of the organizations that participated in the standardization of the CMQ had established compensation systems. These systems usually assign a pay range to every job title. Even within pay ranges, increases in pay are more a function of tenure than of work activities. Therefore, workers are paid according to their job title and not according to their work behavior.

If workers with the same job title are performing the job differently, the format of the scale could influence within job variance. A dichotomous format could encourage agreement among the raters and obscuring real differences among positions with the same job title. This obfuscation would decrease within job variance by eliminating accurate information about the positions. Although the polychotomous format could encourage exaggerations or understatements of the ratings, it does permit job incumbents to describe their positions more accurately. However, this increase in within job variance is not related to the criterion. Pay is more a function of the worker's job title than their work behavior.

It is conceivable that there are valid reasons why workers within the same job title describe their positions differently and that the format of the scale may influence these ratings. It is also possible that pay is more a function of job title than worker behavior. These issues are a source of concern and impact on the results of the present study. The use of pay as the sole criterion may <u>not</u> be appropriate for testing the utility of using polychotomous versus dichotomous scale formats for job analysis data. If the within job variance is high due to actual differences in work activities and pay is a more a function of the job title than work behaviors, then scores based on work behaviors will be less that an optimal predictors of pay.

A more appropriate strategy would be to use the ratings of job analysts as the criterion. If these ratings were based on multiple behavioral observations and interviews, job analysts ratings could be an accurate measure of the workers' behavior. This is the procedure advocated by the authors of the PAQ (McCormick et al., 1972). Although the use of job analysts in gathering information about each position is a costly and time-consuming procedure, the method would be less susceptible to criterion contamination than pay.

The present study also found that (for the scales used) factor and latent trait scores estimated from the importance scale were superior to scores estimated from the frequency scale for polychotomous and dichotomous data. Although it was only possible to test one scale, "Using Information", these results were contrary to what was predicted. It was believed that worker's would be able to more accurately describe their work behavior in terms of frequency rather than importance. However, the results suggest that the relative importance scale provides useful information about the job and that this information is related to pay. Information about the importance of each work behavior may reveal certain aspects of the position that are not captured with a frequency scale. According to Harvey (personal communication, March 16, 1991) the latest version of the CMQ asks workers to rate all items on frequency and importance scales. Further research using the CMQ is needed to assess the value of using importance scale ratings for predicting pay.

It was predicted that number endorsed scoring would be superior to factor scores for predicting pay. The present study demonstrated that number endorsed scores computed from dichotomous data were better predictors of pay than factor or latent trait scores when computed from polychotomous or dichotomous data. The use of polychotomous data adds error variance or reflects actual differences in work activities that are unrelated to pay (as discussed earlier). These problems reduced the value of the number endorsed scale scores computed from polychotomous data for estimating pay. These results support the hypothesis that weighting dichotomous item raw scores with item factor scoring coefficients, reduces the value of the raw scores because the scoring coefficients are unrelated to the criterion.

Number endorsed scoring not only proved to be the best scoring strategy, but also the simplest. Therefore, the relatively complicated and time consuming procedure of computing factor scores for the CMQ is not only unnecessary but may be counter-productive. The results of the present study support Cohen's (1990) assertion that simple scoring procedures are generally superior to more complicated approaches. This researcher recommends the use of unit weighting of scale scores. Weighting scale scores with information that is unrelated to the criterion may reduce or artificially inflate their predictive value. The idea that simpler is better is also asserted by Hersen and Barlow (1976). These researchers caution against mathematical manipulations of data because it can distort or obscure the individual's performance.

In order to determine whether factor scoring and latent trait scoring are equivalent techniques of the same model, the form and the degree of the relationship between factor scores and latent trait scores was tested. The regression of factor scores on latent trait scores for scales two, four, and five do not support Takane and de Leeuw's (1987) assertion that maximum likelihood estimation (MLE) of the two-parameter normal ogive model and factor scoring yield equivalent results. If factor scores and latent trait scores are equivalent procedures of the same model, then there should be a linear relationship between these two variables. This relationship was only seen for scale scores estimated for scale six. Factor and latent trait scale scores for scales two, four, and five demonstrated a quadratic relationship. Also, this relationship was not limited to extreme values of the factor scores or Θs but occurred over the complete range of scores. Therefore, factor and latent trait scores do not appear to be products of equivalent techniques of the same model. They may be two different approaches to estimating the value of the underlying dimension.

The nonlinear relationship between factor and latent trait scores that was demonstrated in the present study supports the findings of Parker (1991) and Harvey (1989). However, those researchers used a three-parameter logistic model. When latent trait scores from the two- and three-parameter logistic model were compared, latent trait scores estimated from the two-parameter model were superior to those computed from the three-parameter model for predicting pay. An examination of the scatterplots for latent trait scores and pay for each model reveals that the three-parameter model appears to alter the relationship between factor and latent trait scores at extreme values. Further research is needed to assess the differences in using the two- and three-parameter logistic models.

One benefit of the IRT model is that test information functions and standard error of theta can be computed for each scale. This advantage allows test developers to assess the range of accurate measurement and add or delete --------items as needed. The results of the present study demonstrated that the four CMQ scales provided accurate estimates of Θ over a relatively narrow range of Θ. The CMQ items accurately measure jobs around the mean but provide little information about high- or low-scoring jobs.

In order for the CMQ to measure all jobs in all situations, items that measure at lower and higher values of the latent trait need to be added to the scales. If dichotomous scoring is to be used, items that discriminate at lower levels of Θ are needed for all scales. Items with difficulty parameters which are greater than 2.3, 1.9, and 2.8 are needed for scales two, five, and six, respectively.

The problem of a lack of items that measure at high and low levels of Θ is not new to job analysis questionnaires. Harvey (1989) demonstrated that the JEI items have limited use for high- and low-scoring jobs. Many of the workers responding to the JEI either endorsed all or none of the items. Therefore, the JEI items were unable to distinguish among high- or low-scoring jobs.

Another universal job analysis instrument, the PAQ has a similar problem. Cascio (1987) and Harvey (1991) assert that the PAQ has limited use for high-scoring jobs because it measures managerial, executive, and professional jobs with an unacceptably high degree of error.

Universal job analysis instruments must be applicable to all jobs and at all levels. In order to accomplish this, items that measure all levels of the common worker-oriented dimensions are needed. While factor analysis can help identify the common work dimensions and the items that measure each dimension, the test developer needs to assess the usefulness of the items in measuring all levels of the work dimension.

Hulin et al., (1983) described a three-stage approach to test development. The first stage involves writing items that measure a single underlying trait. Each of worker-oriented dimensions should contain items that are written to assess different levels of the trait. Second, an IRT model that best represents the data is chosen. The third stage involves the estimation of the test information functions. This enables the test developer to choose items that will provide accurate estimates of Θ over the range of the Θ that is of interest. If a test is designed to assess the extreme values of a trait (e.g. for admission to gifted programs, identify students who need remediation) then items that discriminate at these levels of Θ should be selected. Since the CMQ is intended to be a universal job analysis system, items that represent all possible levels of the worker-oriented dimension need to be included in order to make accurate estimates of Θ over the entire range of Θ.

The results of the present study demonstrate the value of using IRT in the construction of job analysis instruments. The traditional factor analytic method is not sufficient to guarantee the applicability of a job analysis instrument for all jobs and at all levels. IRT analysis is needed to insure that the test is constructed of items that accurately measure a broad range of job dimensions.

Although the results of the present study suggested that IRT was not the best scoring strategy for the CMQ, it can be a productive tool when used for item analysis and test development.

There are three reasons why the results of this study should be interpreted with caution. Since approximately 71% of the questionnaires were administered anonymously, the job incumbents' ratings could not be verified. This casts some doubt about the reliability of these ratings.

Second, only four the CMQ scales were used. The results of the present study may not generalize to the other scales.

Third, the results of the present study could not be cross-validated due the necessity of using a large sample for factor and latent trait analyses. Replications of the present study are needed before firm conclusions regarding the scoring strategy and response format of job analysis instrument can be made.

In summary, the present study contributes to the research in industrial/organizational psychology in several important ways. It raises some concerns about the use of pay as the sole criterion for evaluating job analysis instruments. It challenges the traditional use of factor scoring for computing scale scores and suggests that simpler may be better. It provides evidence that job incumbents can reliability rate the importance of their work behaviors and that these ratings are related to pay. The present study demonstrates the necessity of employing IRT or similar item analysis procedures for constructing universal job analysis instruments which need to measure a broad range of work dimensions.

References

Anastasi, A. (1976). _Psychological testing_ (4th ed.). New York: Macmillan.

Bendig, A. W. (1954). Reliability and the number of rating scale categories. _Journal of Applied Psychology_, _38_, 38-40.

Cascio, W. F. (1987). _Applied psychology in personnel management_. New Jersey: Prentice-Hall.

Cascio, W. F., Valenzi, E. R., & Silbey, V. (1978). Validation and statistical power: Implication for applied research. _Journal of Applied Psychology_, _65_, 35-138.

Cohen, J. (1990). Things I have learned so far. _American Psychologist_, _45_, 1304-1313.

Comrey, A. L., & Montag, I. (1982). Comparison of factor analytic results with two-choice and seven-choice personality item formats. _Applied Psychological Measurement_, _6_, 285-289.

Cornelius, E. T. & Hakel, M. D. (1978). _Job Element Inventory_.

Einhorn, H. J. & Hogarth, R. M. (1975). Unit weighting schemes for decision making. _Organizational Behavior and Human Performance_, _13_, 171-192.

Ferguson, L. W. (1941). A study of the Likert technique of attitude scale construction. _Journal of Social_

Psychology, 13, 51-57.

Gatewood, R. D., & Field, H. S. (1987). Human resource selection. New York: CBS College Publishing.

Green, S. B., & Stutzman, T. (1986). An evaluation of methods to select respondents to structured job-analysis questionnaires. Personnel Psychology, 39, 543-563.

Hambleton, R. & Swaminathan, H. (1985). Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff.

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette (Ed.), Handbook of industrial/organizational psychology (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Harvey, R.J. (1989). Latent trait versus factor scoring of the job element inventory. A paper presented at the American Psychological Association convention, New Orleans.

Hulin, C. L., Drasgow, F. & Parsons, C. K. (1983). Item Response Theory: Application to psychological measurement. Homewood, Illinois: Dow Jones-Irwin.

Jahoda, M,. Deutsch, M., & Cook, S. W. (Eds.) (1951). Research methods in social relations. New York: Dryden Press.

Joe, V. C., & Jahn, C. J. (1973). Factor structure of the rotter I-E scale. Journal of Clinical Psychology, 29, 66-68.

Kaiser, H. F. & Michael, W. B. (1977). Little jiffy factor scores and domain validates. Educational and Psychological Measurement, 37, 363-365.

Lawley, D. N. (1943). On problems connected with item selection and test construction. Proceedings of the Royal Society of England, 61, 273-287.

Lawshe, C.H. & Schucker, R. E. (1959). The relative efficiency of four tests weighting methods in multiple prediction. Educational and Psychological Measurement, 19, 103-114.

Lissitz, R. . W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A monte carlo approach. Journal of Applied Psychology, 60, 10-13.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, N. J.: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley Publishing.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for a likert scale? Study I: Reliability and validity. Educational and Psychological Measurement, 31, 657-674.

McCormick, E.J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology, 56, 347-368.

McCormick, E.J., Jeanneret, P. R., & Mecham, R. C. (1972). Technical manual for the Position Analysis Questionnaire (PAQ). West Lafayette, IN: PAQ Services.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115- 132.

Parker, S. B. (1991). Factor versus latent trait scoring of the Executive Checklist. Paper presented at the meeting of Society of Industrial/Organizational Psychologists, St. Louis, MS.

Rummel, R.J. (1970). Applied Factor Analysis. Evanston, Il: Northwestern University Press.

Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. Psychometric Monograph, No. 17.

Stutzman, T. M. (1983). Within classification job differences. Personnel Psychology, 36, 503-516.

Takane, Y., & Leeuw, J. D. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Thissen, D. & Steinberg, L. (1984). A response model for multiple choice items. Psychometrika, 49, 501-519.

Thissen, D. & Steinberg, L. (1988). Data analysis using item response theory. Psychological Bulletin, 104, 385-395.

Torgerson, W.S. (1958). Theory and method scaling. New York: John Wiley and Sons.

Tucker, L.R. (1946). Maximum validity of a test with equivalent items. Psychometrika, 11, 1-13.

Velicer, W. F., & Stevenson, J. F. (1978). The relation between item format and the structure of the Eysenck Personality Inventory. Applied Psychological Measurement, 2, 293-304.

Wainer, H. (1978). On the sensitivity of regression and regressors. Psychological Bulletin, 85, 267-273.

Wallace, M. J., & Fay, C. H. (1988). Compensation Theory and Practice. Boston: PWS-Kent Publishing.

Wilson, M. A., & Harvey, R. J. (1990). The value of relative-time spent ratings in task-oriented job analysis. Journal of Business and Psychology, 4, 453-461.

Table 1

Proportion of the Variance Accounted for by the First Factor for Polychotomous and Dichotomous Responses to the Frequency and Importance Scales.

Frequency Scale

Factor	Polychotomous Data	Dichotomous Data
One	.8270	.8237
Two	.7620	.7783
Three	.7224	.7544
Four	.7835	.7212
Five	.8594	.8288
Six	.7821	.8522
Seven	.6414	.6703

Importance Scale

Factor	Polychotomous Data	Dichotomous Data
Seven	.6219	.6681

Table 2

Correlation Matrix of Factor Scores (F) and Pay for Scales Two, Four, Five, and Six Estimated from Polychotomous (P) and Dichotomous (D) Data.

	Pay	F2	F4	F5	F6
Pay	1.0	.44	.31	.23	.21
F2		1.0	.48	.25	.41
F4			1.0	.13	.24
F5				1.0	.16
F6					1.0
F2D	.48	.94	.47	.27	.43
F4D	.40	.53	.89	.13	.31
F5D	.23	.26	.14	.94	.16
F6D	.24	.41	.23	.18	.95

	F2D	F4D	F5D	F6D
F2D	1.0	.54	.28	.46
F4D		1.0	.13	.30
F5D			1.0	.18
F6D				1.0

Note. All correlations are significantly different from 0 ($p<.01$).

Table 3

Descriptive Statistics for Pay and Factor Scores (F) for Scales Two, Four, Five, and Six Estimated from Polychotomous and Dichotomous Data.

Frequency Scale
Polychotomous Data

Variable	M	SD	Min	Max
Pay	31879	12771	8160	100000
F2	0	.9743	-.7104	5.1276
F4	0	.9089	-.3882	13.8209
F5	0	.8474	-1.2147	3.4300
F6	0	.9363	-.9032	4.3099

Frequency Scale
Dichotomous Data

Variable	M	SD	Min	Max
F2	0	.9719	-2.7843	2.9993
F4	0	.9232	-.4279	7.1227
F5	0	.8652	-1.3126	2.0066
F6	0	.9462	-.9668	2.2554

Table 4

CMQ Items for Each Scale

Item	Description
Scale Two	
	Attend Meetings Initiated by Other People.
	In order to perform your job, do you attend meetings to...
47	Consult of give specialized information?
48	Informally exchange information or ideas?
49	Formally exchange information or ideas?
50	Coordinate or schedule work activities?
51	Train, instruct, or educate?
52	Supervise or evaluate projects or people?
53	Persuade or sell?
54	Formally bargain or negotiate?
55	Resolve conflicts or disputes?
56	Evaluate options or make a decision?
57	Diagnose or solve problems?
58	Set policies, rules, or procedures?
	Chair or Initiating Meetings.
	In order to perform your job, do you chair or initiate meetings to...
59	Consult or give specialized information?
60	Informally exchange information or ideas?
61	Formally exchange information or ideas?
62	Coordinate or schedule work activities?
63	Train, instruct, or educate?
64	Supervise or evaluate projects or people?
65	Persuade or sell?
66	Formally bargain or negotiate?
67	Resolve conflicts or disputes?
68	Evaluate options or make a decision?
69	Diagnose or solve problems?
70	Set policies, rules, or procedures?

Table 4 (cont'd)

Item	Description
Scale Four	
	Using Information.
	As a source of information that you rely upon to perform your job, do you...
89	Use written words or English?
	Managing Financial Resources
	Are you involved in the making of decisions about...
109	Setting or changing the size of budgets?
112	Managing investments or cash flow?
114	Increasing or decreasing employee salaries or benefits?
115	Establishing or changing the lines of authority and supervision
	Setting long-term business strategies
	Are you involved in the making of decisions about...
123	Taking on a new project?
124	Adding a new product or product line?
125	Increasing the types or levels of services offered to customers or clients?
126	Shutting down or phasing out segments of current operations?
127	Discontinuing products?
128	Discontinuing services?
129	Closing down or abandoning projects?
130	Acquiring business?
131	Starting up new businesses?
132	Selling businesses or subsidiaries?
Scale Five	
	Using Information.
	As a source of information that you rely upon to perform your job, do you...
86	Use spoken words in English?

Table 4 (cont'd)

Item	Description
87	Use written words in English?
90	Use numbers?
91	Use percentages, fractions, decimals?
92	Use algebra and basic statistics?
93	Use higher math (trigonometry, calculus)?
94	Use pictures, drawings, patterns?
95	Use displays, gauges, meters, measuring instruments?
	Using the five senses.
	In order to perform your job, do you use the sense of ...
102	Sight to see differences in colors, patterns, or shapes?
Scale Six	
	People who do not have supervisory job duties.
	Within your organization, do you contact...
21	Unskilled or semi-skilled laborers?
	People who have supervisory job duties.
	Within your organization, do you contact...
22	Unskilled or semi-skilled laborers?
23	Worker directly involved in machine operations, manufacturing, production, or processing?
24	Union stewards or representatives?
25	Personal services employees (waiters, barbers, maids)?
26	Clerical or support staff?
27	Marketing or sales employees?
28	Technical specialists (engineers, programmers, designers)?
29	professional employees lawyers, scientists, doctors)?
30	First-line supervisors?
31	Mid-level managers (department or area managers)?

Table 4 (cont'd)

Item	Description
32	Upper-level managers (regional managers, V.P.'s)?
33	Executives (senior V.P.'s, city managers, C.E.O.'s)?
91	Use percentages, fractions, decimals?
92	Use algebra and basic statistics?
93	Use higher math (trigonometry, calculus)?
94	Use pictures, drawings, patterns?
95	Use displays, gauges, meters, measuring instruments?
96	Observe the quantity or quality of materials or supplies?
97	Observe the physical qualities of people or animals?
98	Observe the behavior and actions of people or animals?
99	Observe the operation and performance of machines or equipment?

Table 5

Summary of Linear and Nonlinear Regressions of Pay on Factor Scores (F) Computed from Polychotomous and Dichotomous Data

Polychotomous Data

Source	DF	F
Model	4	206.0**
Error	2683	

R-Square=.2352

Variable	PE	SE	F
F2	6436.01	404.65	252.97**
F4	2026.81	271.84	55.59**
F5	1650.01	264.23	39.00**
$F2^2$	-1290.38	165.77	60.59**

Dichotomous Data

Source	DF	F
Model	3	323.39**
Error	2683	

R-Square=.2658

Variable	PE	SE	F
F2	4392.63	267.05	270.55**
F4	2784.21	272.72	104.23**
F5	1692.23	254.22	44.31**

Note. **p<.01

Table 6

Descriptive Statistics for Absolute Residuals from Regressions of Pay on Factor Scores (F) Computed from Frequency Scale Polychotomous (P) and Dichotomous (D) Data.

Scales Two, Four, Five, and Six

Data	*M*	*SD*	Max	Min
P	8676.12	7031.38	72170.36	.40
D	8531.02	6852.02	72036.31	1.91

Table 7

Summary of t-tests of Mean Absolute Residuals of Pay from the Regression of Pay on Factor Scores from Polychotomous and Dichotomous Data

Scores	*M*	Max	Min	*t*
F	1248.56	14972.08	.55	41.81**

$^{**}p<.01$.

Table 8

Descriptive Statistics for Latent Trait Scores (LT) Estimated from the Frequency Scale and Polychotomous and Dichotomous Data

Polychotomous Data

Variable	M	SD	Min	Max
LT 2	-2.3685	.9980	-3.4090	1.8960
LT 4	-2.7528	.4205	-2.9970	.7330
LT 5	-1.2086	.8046	-2.5420	1.1900
LT 6	-1.8899	.9586	-2.8710	2.8710

Dichotomous Data

Variable	M	SD	Min	Max
LT 2	-2.6152	.6786	-3.4090	-1.0590
LT 4	-2.7869	.3509	-2.9970	-1.0830
LT 5	-1.7710	.4252	-2.5420	-.9320
LT 6	-2.2562	.5569	-2.8710	-.9910

Table 9

Summary of Linear and Nonlinear Regressions of Pay on Latent Trait Scores (LT) Computed from Polychotomous and Dichotomous Data

Polychotomous Data

Source	DF	F
Model	3	301.39**
Error	2681	
R-Square=.2522		

Variable	PE	SE	F
LT2	4236.10	264.13	257.21**
LT4	5718.79	606.33	88.96**
LT5	1744.84	276.99	39.68**

Dichotomous Data

Source	DF	F
Model	3	324.69**
Error	2681	
R-Square=.2666		

Variable	PE	SE	F
LT2	5950.89	388.20	234.99**
LT4	7797.88	719.78	117.37**
LT5	3475.69	526.72	43.54**

Note. $**p<.01$.

Table 10

Descriptive Statistics for Absolute Residuals from Regressions of Pay on Latent Trait Scores Computed from the Frequency Scale and Polychotomous (P) and Dichotomous (D) Data.

Scales Two, Four, Five, and Six

Data	M	SD	Max	Min
P	8610.71	6913.13	71983.41	7.49
D	8548.01	6821.66	71845.25	1.99

Table 11

Summary of Paired Comparisons of Mean Absolute Residuals of Pay from the Regression of Pay on Latent Trait Scores Estimated from Polychotomous Versus Dichotomous Data

Scores	*M*	Min	Max	*t*
LT	736.71	.61	11301.15	48.11**

Note. **$p < .01$

Table 12

CMQ Items for Scale Seven

Item	Description
	Using Information As a source of information that you rely upon to perform your job, do you...
86	Use spoken words in English?
87	Use written works in English?
88	Use spoken words in a foreign language?
89	Use written words in a foreign language?
90	Use numbers?
91	Use percentages, fractions, decimals?
92	Use algebra and basic statistics?
93	Use higher math (trigonometry, calculus)?
94	Use pictures, drawings, patterns?
95	Use displays, gauges, meters, measuring instruments?
96	Observe the quantity or quality of materials or supplies?
97	Observe the physical qualities of people or animals?
98	Observe the behavior and actions of people or animals?
99	Observe the operation and performance of machines or equipment?

Table 13

Descriptive Statistics of Factor Scores for Scale Seven Estimated from the Frequency and Importance Scales and Polychotomous and Dichotomous Data

Frequency Scale

Data	*M*	SD	Min	Max
P	0	.8991	-2.1113	2.7831
D	0	.9003	-1.9339	2.0814

Importance Scale

Data	*M*	SD	Min	Max
P	0	.8858	-1.3601	5.0327
D	0	.9021	-2.1628	1.9364

Table 14

Summary of Linear and Nonlinear Regressions of Pay on Factor Scores (F), Scale Seven, Computed from the Frequency and Importance Scales and Polychotomous and Dichotomous Data

Frequency Scale
Polychotomous Data

Source	DF		F
Model	3		122.75**
Error	2683		
R-Square=.0839			
Variable	PE	SE	F
F 7	6779.24	498.90	184.64**
$F7^3$	-1347.88	180.92	55.50**

Importance Scale
Polychotomous Data

Source	DF		F
Model	3		131.16**
Error	2681		
R-Square=.0891			
Variable	PE	SE	F
F7	5214.75	336.68	239.91**
$F7^3$	-359.69	62.30	33.33**

Table 14 (cont'd)

Frequency Scale
Dichotomous Data

Source	DF	F
Model	3	204.07**
Error	2681	
R-Squared=.1321		

Variable	PE	SE	F
F7	5207.74	262.77	392.78**
$F7^4$	629.45	73.87	72.60**

Importance Scale
Dichotomous Data

Source	DF	F
Model	3	205.81**
Error	2681	
R-Squared=.1331		

Variable	PE	SE	F
F7	5956.36	294.33	409.54**
$F7^4$	640.48	56.15	130.10**

Note. $**p<.01$

Table 15

Descriptive Statistics for Absolute Residuals from Regressions of Pay on Factor Scores, Scale Seven, Estimated from Frequency and Importance Scales and Polychotomous and Dichotomous Data.

Polychotomous Data

Scale	M	SD	Max	Min
Frequency	8676.12	7031.38	72170.36	.40
Importance	8610.71	6913.13	71983.41	7.49

Dichotomous Data

Scale	M	SD	Max	Min
Frequency	8531.02	6852.22	72063.31	1.91
Importance	8548.01	6821.66	71845.25	1.99

Table 16

Summary of t-Tests of Mean Absolute Residuals of Pay from the Regression of Pay on Factor Scores from Frequency Versus Importance Scales and Polychotomous (P) and Dichotomous (D) Data

Data	*M*	Min	Max	*t*
P	2493.96	4.89	20905.62	66.23**
D	760.81	.05	21506.36	30.27**

Note. **$p<.01$.

Table 17

Descriptive Statistics for Latent Trait Scores, Scale Seven, Estimated from Frequency (F) and Importance (I) Scales and Polychotomous (P) and Dichotomous (D) Data

Scale	Data	*M*	*SD*	Min	Max
F	P	-.9217	.9704	-2.9360	2.9360
F	D	-.2780	.6873	-1.8700	1.9700
I	P	-1.6808	.5602	-2.9360	.8820
I	D	-.2663	.6923	-1.8700	1.8700

Table 18

Correlation Matrix of Latent Trait Scores, Scale Seven, Estimated from Frequency (F) and Importance (I) Scales and Polychotomous (P) and Dichotomous (D) Data

	Pay	FP	IP	FD	ID
Pay	1.0	.28	.29	.31	.30
FP		1.0	.74	.96	.84
IP			1.0	.80	.93
FD				1.0	.87
ID					1.0

Note. All correlation are significantly different from 0 ($p<.01$).

Table 19

Summary of Linear and Nonlinear Regressions of Pay on Latent Trait Scores (LT), Scale Seven, Computed from Frequency and Importance Scales and Polychotomous and Dichotomous Data

Frequency Scale
Polychotomous Data

Source	DF		F
Model	3		06.51**
Error	2680		
R-Square=.0872			
Variable	PE	SE	F
LT 7	4867.13	439.72	122.52**
LT 7^2	-2060.61	372.60	30.59**
LT 7^3	-1004.69	129.93	59.80**

Importance Scale
Polychotomous Data

Source	DF		F
Model	3		126.43**
Error	2680		
R-Square=.1240			
Variable	PE	SE	F
LT 7	-14510.33	1851.15	20.64**
LT 7^2	-12854.75	3193.56	71.28**
LT 7^4	857.83	1522.60	105.02**

Table 19 (cont'd)

Frequency Scale
Dichotomous Data

Source	DF		F
Model	3		196.87**
Error	2681		
R-Square=.1281			
Variable	PE	SE	F
LT 7	10050.60	553.13	330.17**
LT 7^3	-2524.70	262.29	92.66**

Importance Scale
Dichotomous Data

Source	DF		F
Model	3		205.27**
Error	2681		
R-Square=.1328			
Variable	PE	SE	F
LT 7	33341.96	549.38	363.99**
LT 7^3	10481.46	248.73	123.30**

Note. ** $p<.01$.

Table 20

Descriptive Statistics for Absolute Residuals of Pay from Regressions of Pay on Latent Trait Scores Computed from Frequency and Importance Scales and Polychotomous and Dichotomous Data.

Polychotomous Data

Scale	M	SD	Max	Min
Frequency	9237.66	7863.93	71032.86	1.59
Importance	9012.25	7850.94	68352.03	.03

Dichotomous Data

Scale	M	SD	Max	Min
Frequency	9047.83	7767.75	70821.10	6.11
Importance	8998.79	7774.50	68197.10	6.04

Table 21

Summary of t-Tests of Mean Absolute Residuals of Pay from the Regression of Pay on Latent Trait Scores, Scale Seven, Estimated from Frequency and Importance Scales and Polychotomous (P) and Dichotomous (D) Data

Data	*M*	Min	Max	*t*
P	1996.80	.51	23761.06	62.43**
D	591.06	4.61	15108.99	24.87**

**p<.01.

Table 22

Descriptive Statistics for Numbered Endorsed Scores (N-E) Estimated from the Frequency Scale and Polychotomous and Dichotomous Data

Polychotomous Data

Variable	M	SD	Min	Max
N-E 2	13.5738	16.6547	0	96.00
N-E 4	1.8148	3.6810	0	50.00
N-E 5	11.8443	7.1860	0	35.00
N-E 6	10.2321	10.6079	0	65.00

Dichotomous Data

Variable	M	SD	Min	Max
N-E 2	5.6122	6.0324	0	24.00
N-E 4	1.0313	1.9568	0	14.00
N-E 5	3.4720	2.2515	0	9.00
N-E 6	3.2183	3.2514	0	13.00

Table 23

Correlation Matrix of Number Endorsed Scores (N-E) Estimated from the Frequency Scale and Polychotomous and Dichotomous Data

Polychotomous Data

	Pay	N-E 2	N-E 4	N-E 5	N-E 6
Pay	1.0	.44	.32	.20	.20
N-E 2		1.0	.51	.24	.41
N-E 4			1.0	.10	.24
N-E 5				1.0	.18
N-E 6					1.0

Dichotomous Data

	Pay	N-E 2	N-E 4	N-E 5	N-E 6
Pay	1.0	.48	.41	.27	.24
N-E 2		1.0	.57	.33	.48
N-E 4			1.0	.18	.33
N-E 5				1.0	.22
N-E 6					1.0

Note. All correlation are significantly different from 0 ($p<.01$).

Table 24

Summary of Linear and Nonlinear Regressions of Pay on Number Endorsed Scores (N-E) Computed from Polychotomous and Dichotomous Data

Polychotomous Data

Source	DF	F
Model	3	258.85**
Error	2683	
R-Square=.2247		

Variable	PE	SE	F
N-E 2	511.78	328.19	225.78**
N-E 4	504.08	69.11	53.21**
N-E 2^2	-5.08	.56	52.44**

Dichotomous Data

Source	DF	F
Model	3	333.02**
Error	2683	
R-Square=.2716		

Variable	PE	SE	F
N-E 2	682.64	44.36	236.80**
N-E 4	1323.26	131.11	101.86**
N-E 5	715.03	99.11	52.04**

Note. $^{**}p<.01$.

Table 25

Descriptive Statistics for Absolute Residuals of Pay from Regressions of Pay on Number Endorsed Scores Computed from Polychotomous (P) and Dichotomous (D) Data.

Data	*M*	SD	Max	Min
P	8840.84	6948.16	72120.46	5.31
D	8501.64	6820.05	2321.83	.60

Table 26

Summary of t-Tests of Mean Absolute Residuals of Pay from the Regression of Pay on Number Endorsed, Factor, and Latent Trait Scores from Frequency Scale and Polychotomous and Dichotomous Data

Number Endorsed Scores
Polychotomous Versus Dichotomous Data

M	*Min*	*Max*	*t*
1807.09	1.68	12380.39	58.49**

Factor Scores Versus Number Endorsed Scores
Polychotomous Data

M	*Min*	*Max*	*t*
1191.61	0.15	9128.71	65.24**

Factor Scores Versus Number Endorsed Scores
Dichotomous Data

M	*Min*	*Max*	*t*
647.77	0.44	3914.19	64.19**

Latent Trait Scores Versus Factor Scores
Polychotomous Data

M	*Min*	*Max*	*t*
1411.36	0.50	10574.34	66.30**

Latent Trait Scores Versus Factor Scores
Dichotomous Data

M	*Min*	*Max*	*t*
848.92	3.54	7876.10	61.78**

Note. $^{**}p<.01$.

Table 27

Correlation Matrix of Factor and Latent Trait Scores Estimated from Polychotomous and Dichotomous Data

Polychotomous Data

	F2	F4	F5	F6	LT2	LT4	LT5	LT6
F2	1.0	.48	.25	.41	.95	.56	.25	.41
F4		1.0	.13	.24	.48	.92	.12	.23
F5			1.0	.16	.28	.14	.97	.19
F6				1.0	.45	.30	.17	.98
LT2					1.0	.54	.29	.46
LT4						1.0	.14	.30
LT5							1.0	.19
LT6								1.0

Dichotomous Data

	F2	F4	F5	F6	LT2	LT4	LT5	LT6
F2	1.0	.54	.28	.46	.95	.58	.30	.46
F4		1.0	.13	.30	.49	.95	.16	.30
F5			1.0	.18	.30	.15	.98	.19
F6				1.0	.50	.33	.21	.99
LT2					1.0	.55	.33	.50
LT4						1.0	.18	.33
LT5							1.0	.21
LT6								1.0

Note. All correlations are significantly different from 0 ($p<.01$).

Table 28

Summary of Linear and Nonlinear Regressions of Factor Scores (F) on Latent Trait Scores (LT) Computed from Polychotomous and Dichotomous Data

Polychotomous Data
Scale Two

Source	DF	F	
Model	2	50361.6**	
Error	2681		
R-Square=.9741			
Variable	PE	SE	F
LT 2	1.5082	0.0077	38511.3**
LT 2^3	- 0.0443	0.0005	6684.17**

Polychotomous Data
Scale Four

Source	DF	F	
Model	2	17942.60**	
Error	2681		
R-Square=.9305			
Variable	PE	SE	F
LT 4	5.6828	0.0657	7479.40**
LT 4^2	0.8151	0.0143	3252.80**

Table 28 (cont'd)

Polychotomous Data
Scale Five

Source	DF	F
Model	2	46351.9**
Error	2681	
R-Square=.9719		

Variable	PE	SE	F
LT 5	1.4356	0.0098	21340.80**
LT 5^2	0.1745	0.0039	1962.70**

Polychotomous Data
Scale Six

Source	DF	F
Model	1	69484.50**
Error	2682	
R-Square=.9628		

Variable	PE	SE	F
LT 6	1.8113	0.0077	55269.90**

Dichotomous Data
Scale Two

Source	DF	F
Model	2	88365.30**
Error	2681	
R-Square=.9851		

Variable	PE	SE	F
LT 2	4.5575	0.0261	30390.00**
LT 2^2	0.6611	0.0053	15224.10**

Dichotomous Data
Scale Four

Source	DF	F
Model	2	74473.80**
Error	2681	
R-Square=.9822		

Variable	PE	SE	F
LT 4	9.7798	0.0672	21184.20**
LT 4^2	1.4944	0.0137	11852.00**

Dichotomous Data
Scale Five

Source	DF	F
Model	2	77732.90**
Error	2681	
R-Square=.9830		

Variable	PE	SE	F
LT 5	4.1341	0.0409	10194.70**
LT 5^2	0.5936	0.0113	2760.64**

Table 28 (cont'd)

Dichotomous Data
Scale Six

Source	DF	*F*
Model	1	110036.00**
Error	2682	
R-Square=.9762		

Variable	PE	SE	*F*
LT 6	1.6788	0.0051	103719.00**

Note. **$p<.01$.

Table 29

Descriptive Statistics for Latent Trait Scores (LT) Estimated with the Three-Parameter Model and Dichotomous Data

Variable	M	SD	Min	Max
LT 2	-.8507	.7206	-1.5920	1.5400
LT 4	-1.2847	.3609	-1.4790	1.0210
LT 5	-.3793	.5737	-1.2830	1.2320
LT 6	-.7716	.6410	-1.424	1.3720

Table 30

Correlation Matrix of Factor Scores (F) and Latent Trait Scores (LT) Estimated with the Three-Parameter Model and Dichotomous Data

	F2	F4	F5	F6	LT2	LT4	LT5	LT6
F2	1.0	.54	.28	.46	.95	.57	.31	.46
F4		1.0	.13	.30	.22	.98	.17	.30
F5			1.0	.18	.30	.10	.99	.19
F6				1.0	.49	.32	.21	.99
LT2					1.0	.57	.34	.49
LT4						1.0	.18	.32
LT5							1.0	.22
LT6								1.0

Note. All correlations are significantly different form 0 ($p<.01$).

Table 31

Summary of Linear and Nonlinear Regressions of Factor Scores on Latent Trait Scores (LT) Estimated with the Three-Parameter Model and Dichotomous Data

Scale Two

Source	DF	F
Model	2	22266.30**
Error	2681	
R-Square=.9432		

Variable	PE	SE	F
LT 2	1.8316	.0149	15195.90**
LT 2^3	-.2553	.0063	1635.56**

Scale Four

Source	DF	F
Model	2	72791.20**
Error	2681	
R-Square=.9819		

Variable	PE	SE	F
LT 4	3.5519	.0181	38141.40**
LT 4^2	.6423	.0103	59.43**

Table 31 (cont'd)

Scale Five

Source	DF	F
Model	2	11294.00**
Error	2681	

R-Square=.9765

Variable	PE	SE	F
LT 5	1.4904	.0045	11294.00**

Scale Six

Source	DF	F
Model	1	92560.60**
Error	2682	

R-Square=.9718

Variable	PE	SE	F
LT 6	1.4551	.0048	92560.60**

Note. **p<.01.

Table 32

Summary of Linear and Nonlinear Regressions of Pay on Latent Trait Scores (LT) Estimated with the Three-Parameter Model and Dichotomous Data

Source	DF	F
Model	2	347.59**
Error	2681	
R-Square=.2059		

Variable	PE	SE	F
LT 2	13169.70	619.28	452.26**
LT 5	4470.66	389.61	131.67**

Note. **p<.01.

Table 33

Descriptive Statistics for Absolute Residuals of Pay from Regressions of Pay on Latent Trait Scores Estimated with the Three-Parameter Model and Dichotomous Data.

M	SD	Max	Min
15453.40	9976.56	62535.09	35.09

Table 34

Summary of t-Test of Mean Absolute Residuals of Pay from the Regression of Pay on Latent Trait Estimated with the Two- versus Three-Parameter Models and Dichotomous Data

M	Min	Max	*t*
9949.05	9.72	41318.02	72.89**

Note. **$p<.01$.

Table 35

Descriptive Statistics for Latent Trait Parameters Estimated with the Two-Parameter and Graded Response Models.

Two-Parameter Logistic Model

Scale	Parameter	n	M	SD	Min	Max
LT 2	a	24	2.19	.57	1.22	3.35
LT 2	b	24	.99	.71	-.70	2.28
LT 4	a	15	2.34	.57	1.16	3.17
LT 4	b	15	2.22	.93	.81	4.20
LT 5	a	9	1.52	.41	.85	2.02
LT 5	b	9	.56	.96	-1.14	1.92
LT 6	a	13	2.49	1.21	.90	5.04
LT 6	b	13	1.18	.88	.08	2.77

Graded Response Model

Scale	Parameter	n	M	SD	Min	Max
LT 2	a	24	2.10	.48	1.34	3.05
LT 2	b_1	24	.62	.68	-1.05	1.85
LT 2	b_2	24	.87	.65	-.64	2.07
LT 2	b_3	24	1.4	.61	.28	2.58
LT 2	b_4	24	2.24	.58	1.48	3.44
LT 2	b_5	24	3.68	.90	2.59	6.07
LT 4	a	15	2.74	.68	1.61	4.08
LT 4	b_1	15	1.76	.54	.72	2.67
LT 4	b_2	15	2.23	.50	1.15	2.94
LT 4	b_3	15	2.64	.39	1.80	3.26
LT 4	b_4	15	3.11	.61	2.34	4.57
LT 4	b_5	15	3.45	1.22	-.46	5.22

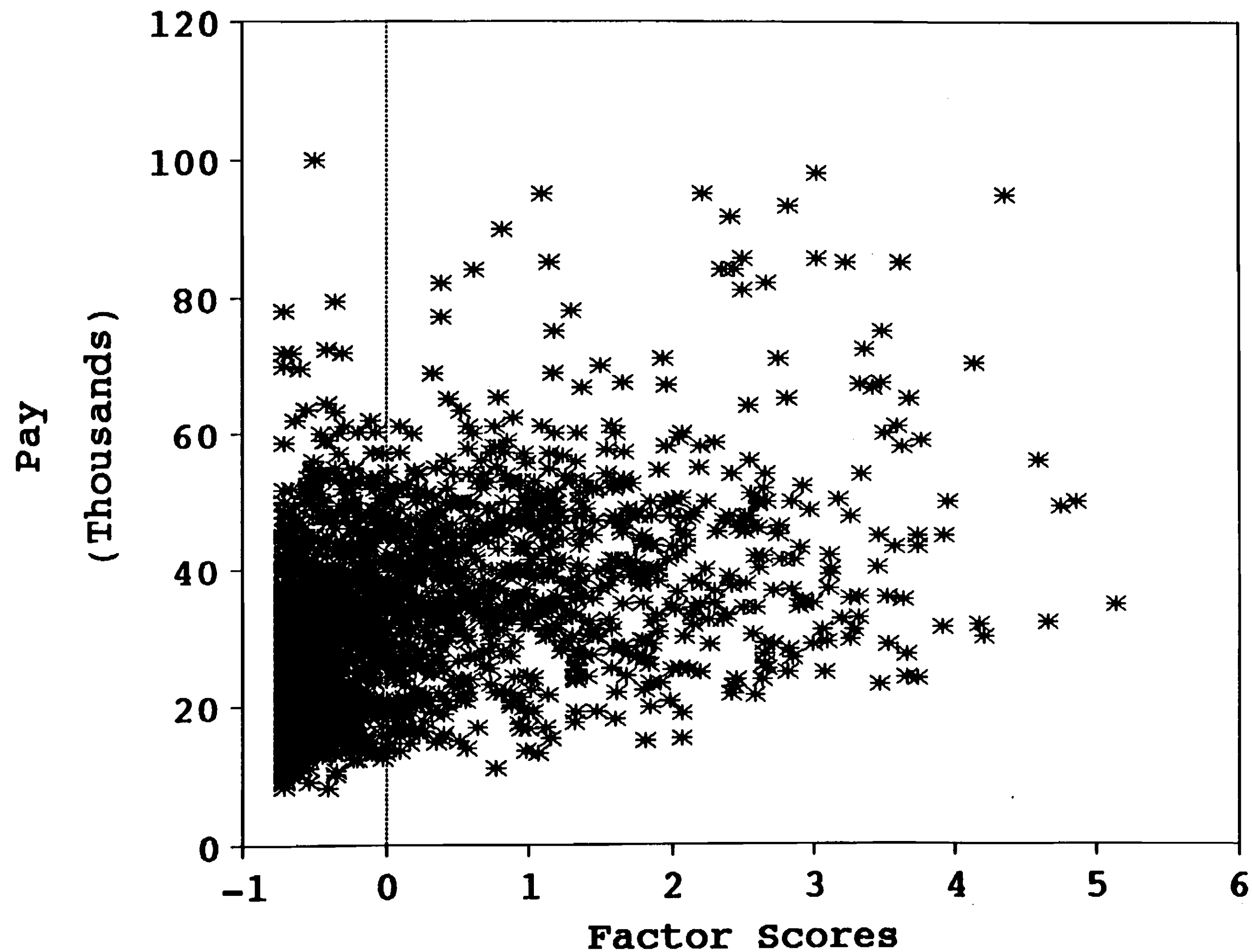
Table 35 (cont'd)

Scale	Parameter	n	M	SD	Min	Max
LT 5	a	7	1.38	.70	.65	2.77
LT 5	b_1	7	.70	.84	-.70	1.89
LT 5	b_2	7	1.32	1.02	-.73	2.60
LT 5	b_3	7	1.91	1.23	-.55	3.37
LT 5	b_4	7	3.26	1.77	-.14	5.20
LT 5	b_5	7	4.40	1.70	1.90	6.78
LT 6	a	13	1.97	.72	.81	3.08
LT 6	b_1	13	.78	.91	-.62	2.34
LT 6	b_2	13	.92	1.00	-.64	2.58
LT 6	b_3	13	1.15	1.03	-.51	2.81
LT 6	b_4	13	1.56	1.03	-.11	3.32
LT 6	b_5	13	2.83	1.18	1.71	5.23

Table 36

Range of Accurate Measurement (95% Confidence Interval) for Latent Trait Scores Estimated from Polychotomous and Dichotomous Data

Scale	Type of Data	Range of Theta	z
LT 2	Polychotomous	-.3 to 1.7	2.0
LT 2	Dichotomous	-.1 to 2.1	2.2
LT 4	Polychotomous	1.0 to 2.7	1.7
LT 4	Dichotomous	1.2 to 2.4 and 2.8 to 3.1	1.5
LT 5	Polychotomous	none	0
LT 5	Dichotomous	none	0
LT 6	Polychotomous	-.1 to .9	1.0
LT 6	Dichotomous	-.1 to 1.2	1.3



<u>Figure 1.</u> Scale 2 (factor scores by pay, frequency, polychotomous data).

Table 37

Reliability Coefficients for Factor and Number Endorsed Scores

Factor Scores

Scale	Type of Scale	Type of Data	Reliability
F2	Frequency	Polychotomous	.8230
F2	Frequency	Dichotomous	.8263
F4	Frequency	Polychotomous	.9474
F4	Frequency	Dichotomous	.9432
F5	Frequency	Polychotomous	.8488
F5	Frequency	Dichotomous	.8715
F6	Frequency	Polychotomous	.8641
F6	Frequency	Dichotomous	.8686
F7	Frequency	Polychotomous	.8295
F7	Frequency	Dichotomous	.8427
F7	Importance	Polychotomous	.8333
F7	Importance	Dichotomous	.8388

Number Endorsed Scores

Scale	Type of Scale	Type of Data	Reliability
N-E 2	Frequency	Polychotomous	.9390
N-E 2	Frequency	Dichotomous	.9334
N-E 4	Frequency	Polychotomous	.7723
N-E 4	Frequency	Dichotomous	.8193
N-E 5	Frequency	Polychotomous	.6730
N-E 5	Frequency	Dichotomous	.7297
N-E 6	Frequency	Polychotomous	.8419
N-E 6	Frequency	Dichotomous	.8624

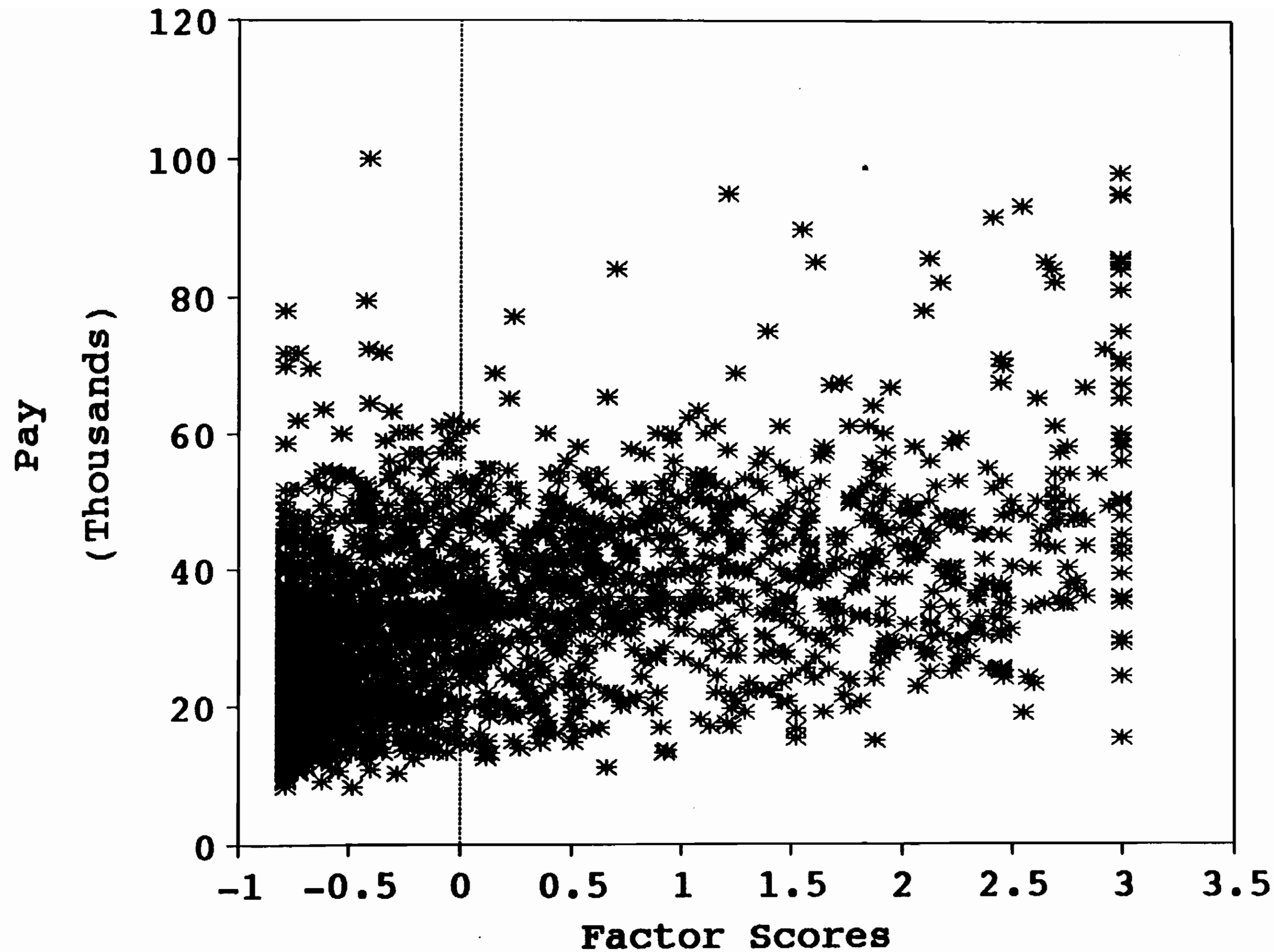


__Figure 2.__ Scale 2 (factor scores by pay, frequency, dichotomous data).

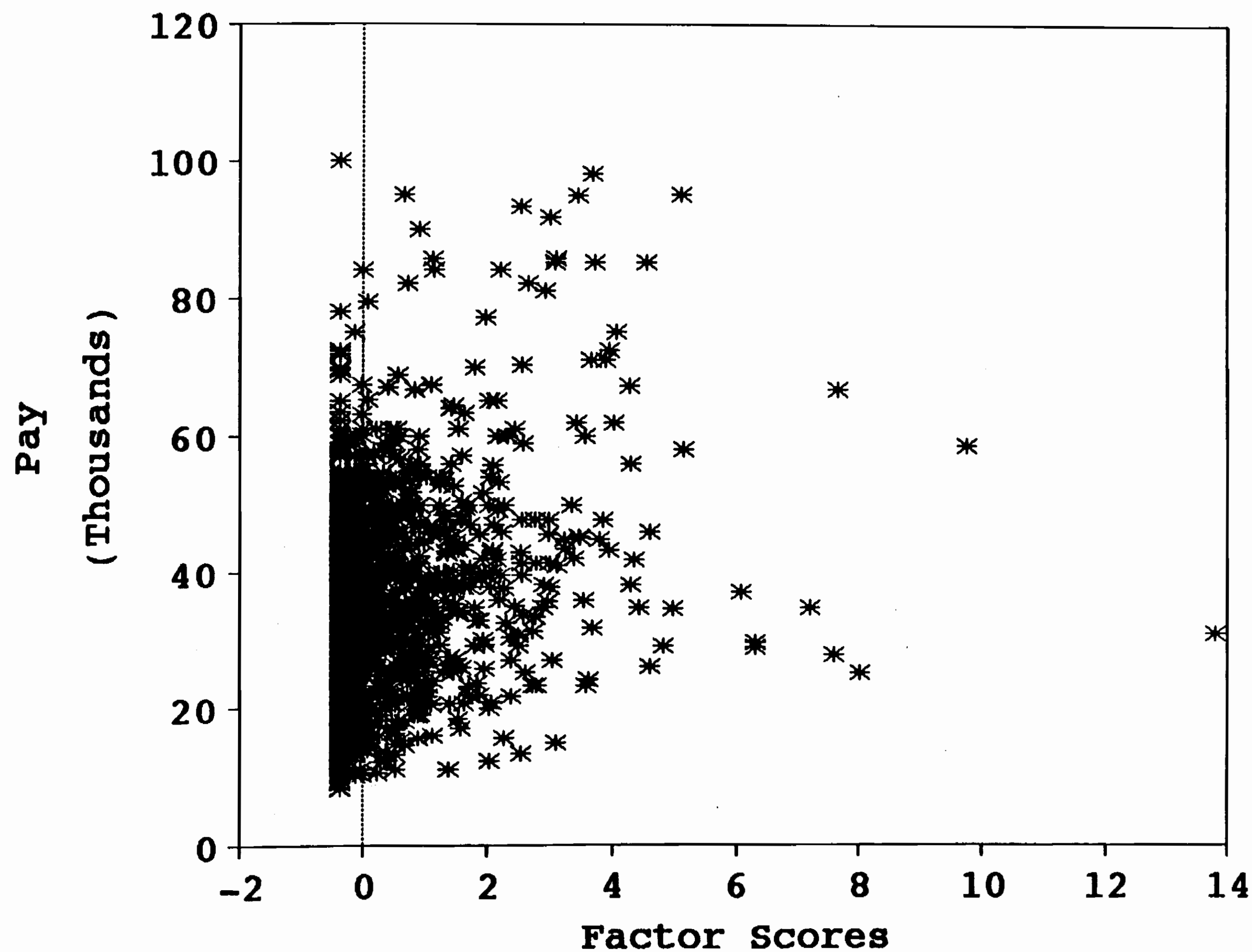


Figure 3. Scale 4 (factor scores by pay, frequency, polychotomous data).

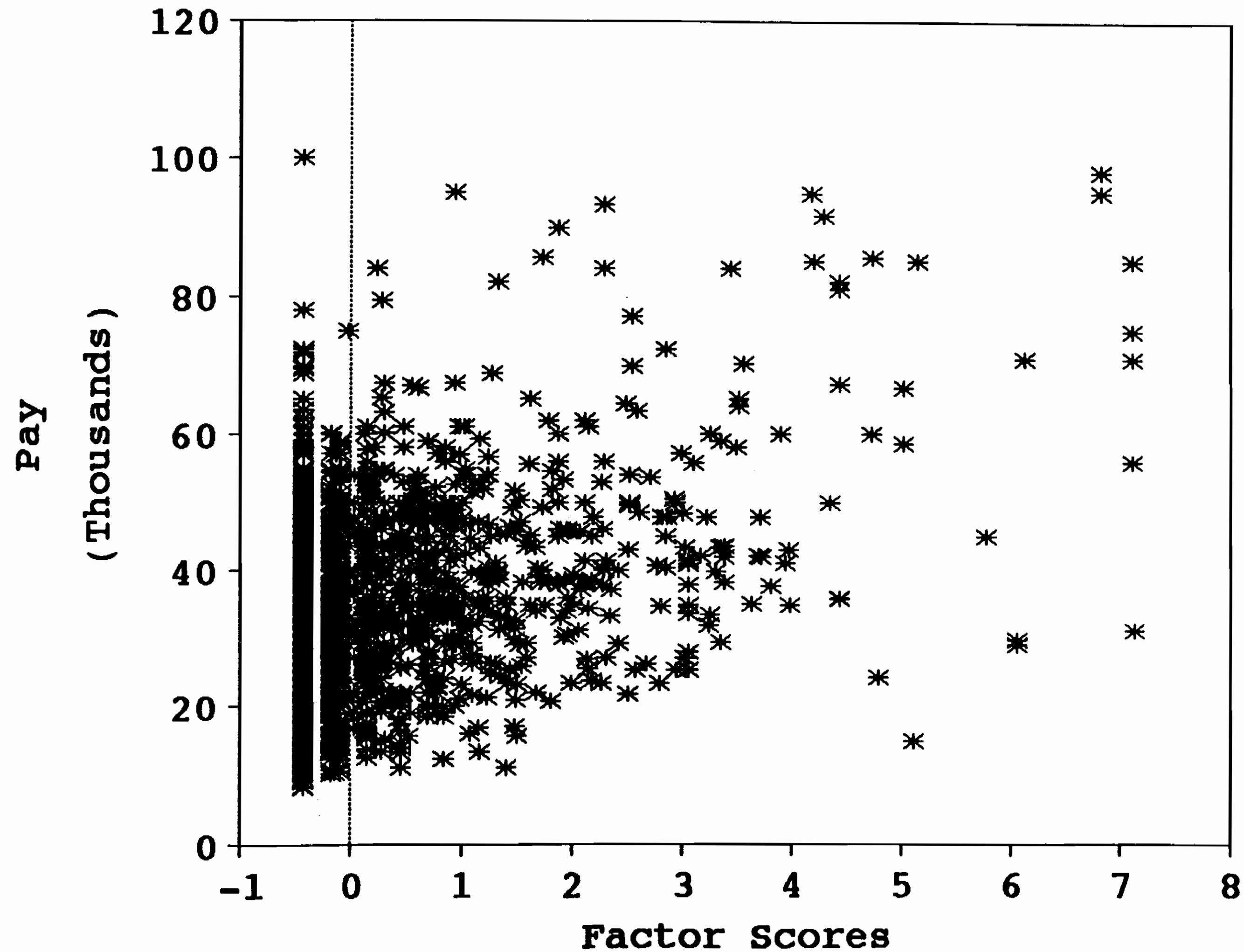


Figure 4. Scale 4 (factor scores by pay, frequency, dichotomous data).

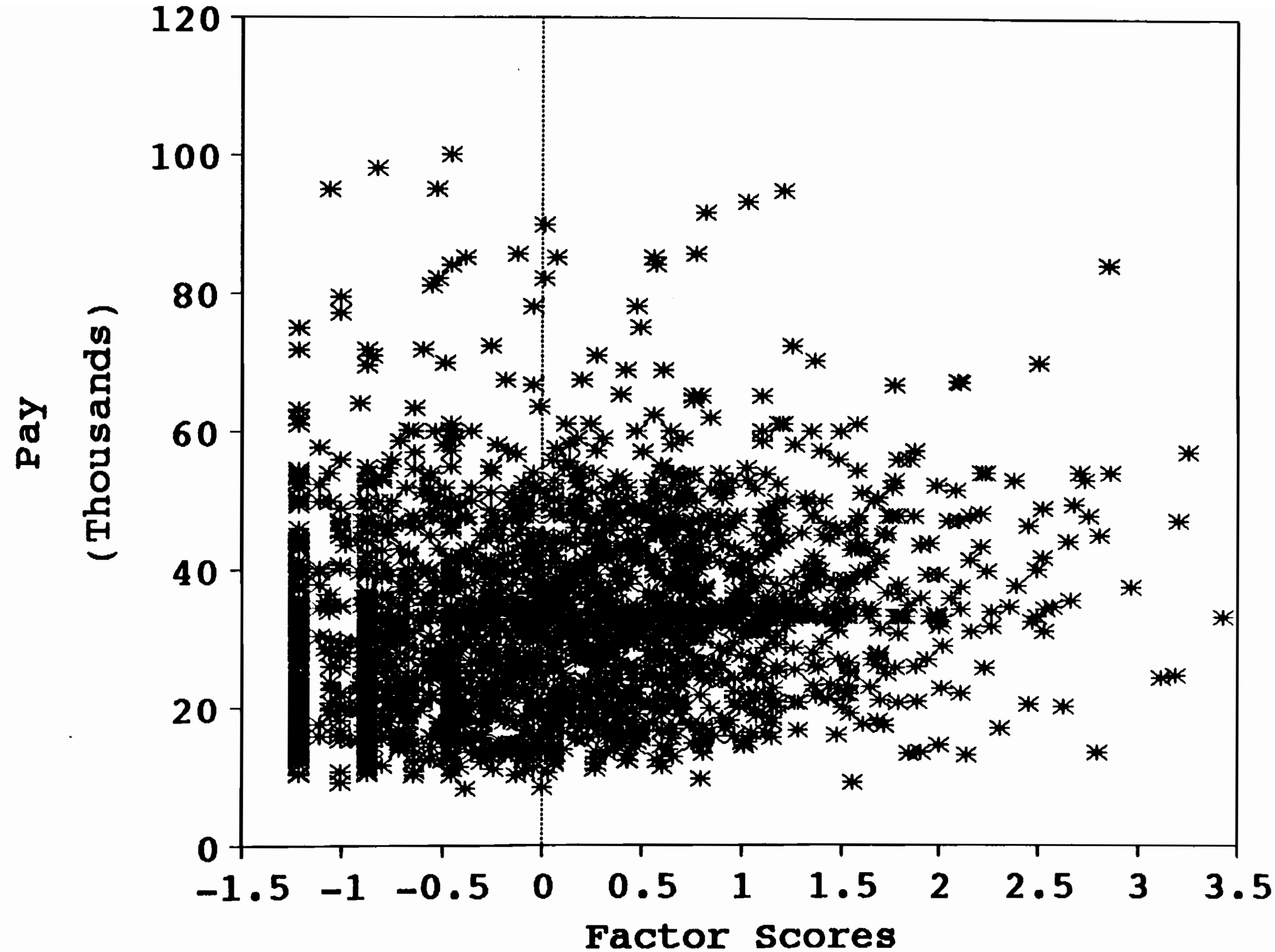
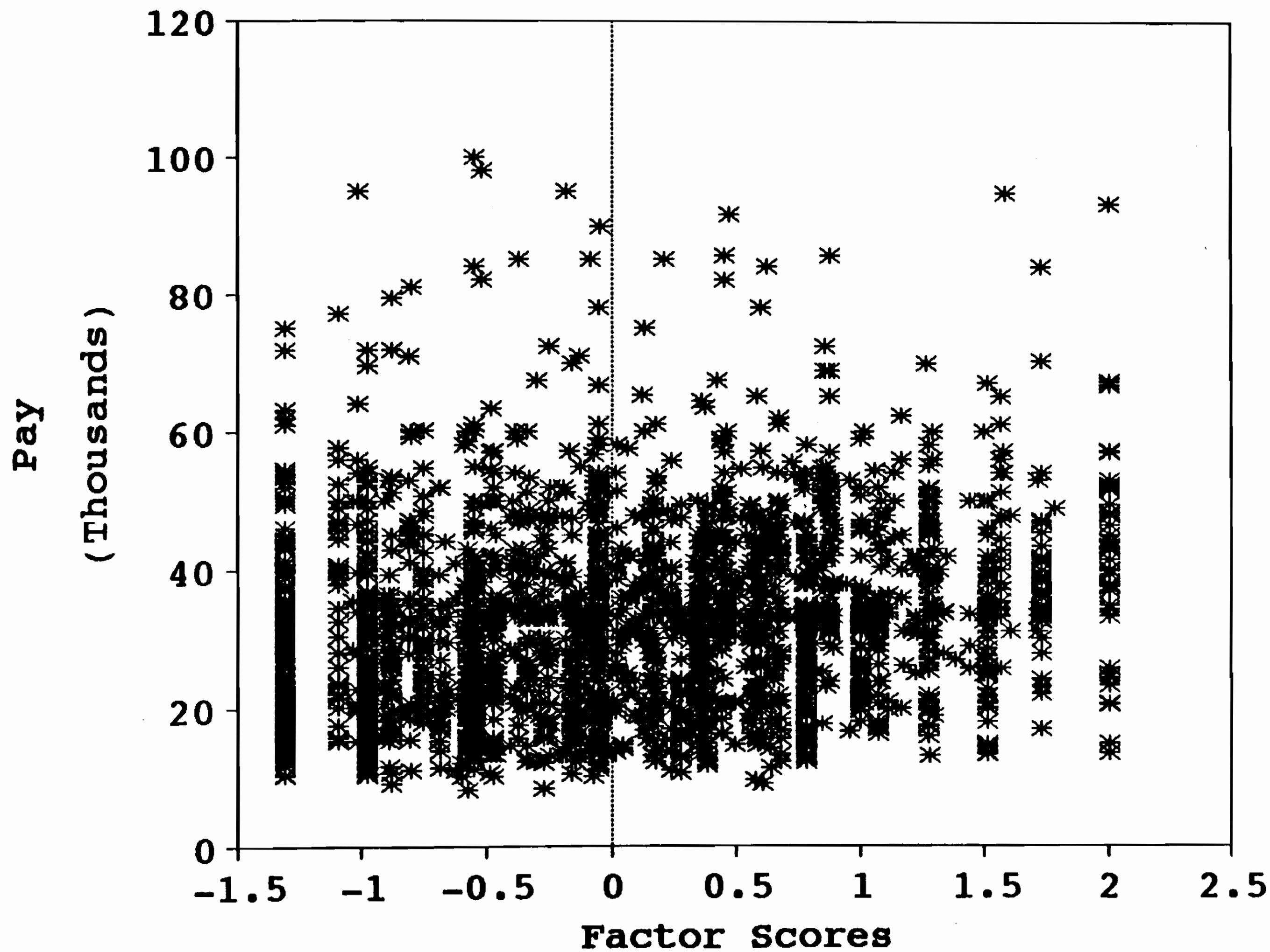
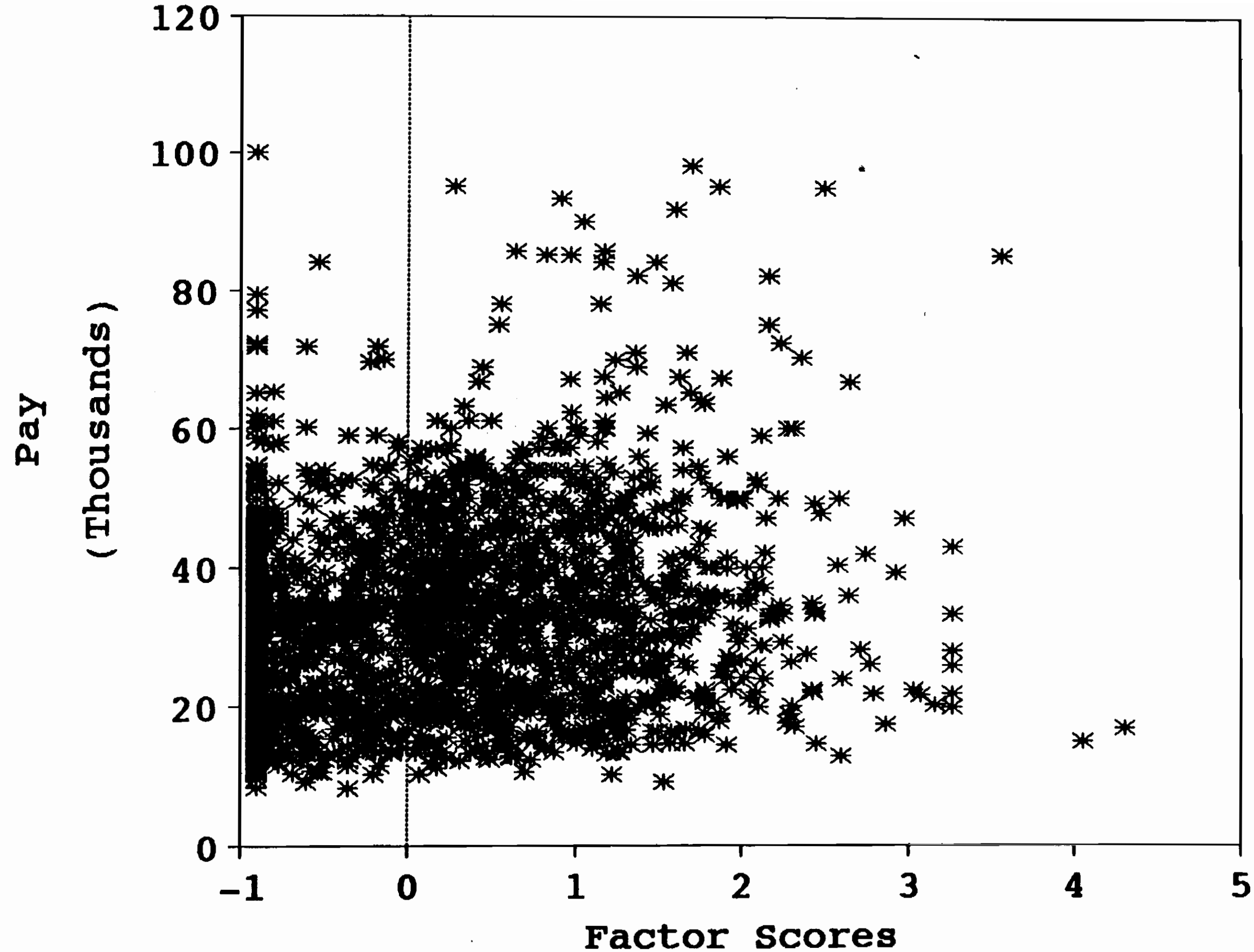


Figure 5. Scale 5 (factor scores by pay, frequency, polychotomous data).



<u>Figure 6.</u> Scale 5 (factor scores by pay, frequency, dichotomous data).



<u>Figure 7.</u> Scale 6 (factor scores by pay, frequency, polychotomous data).

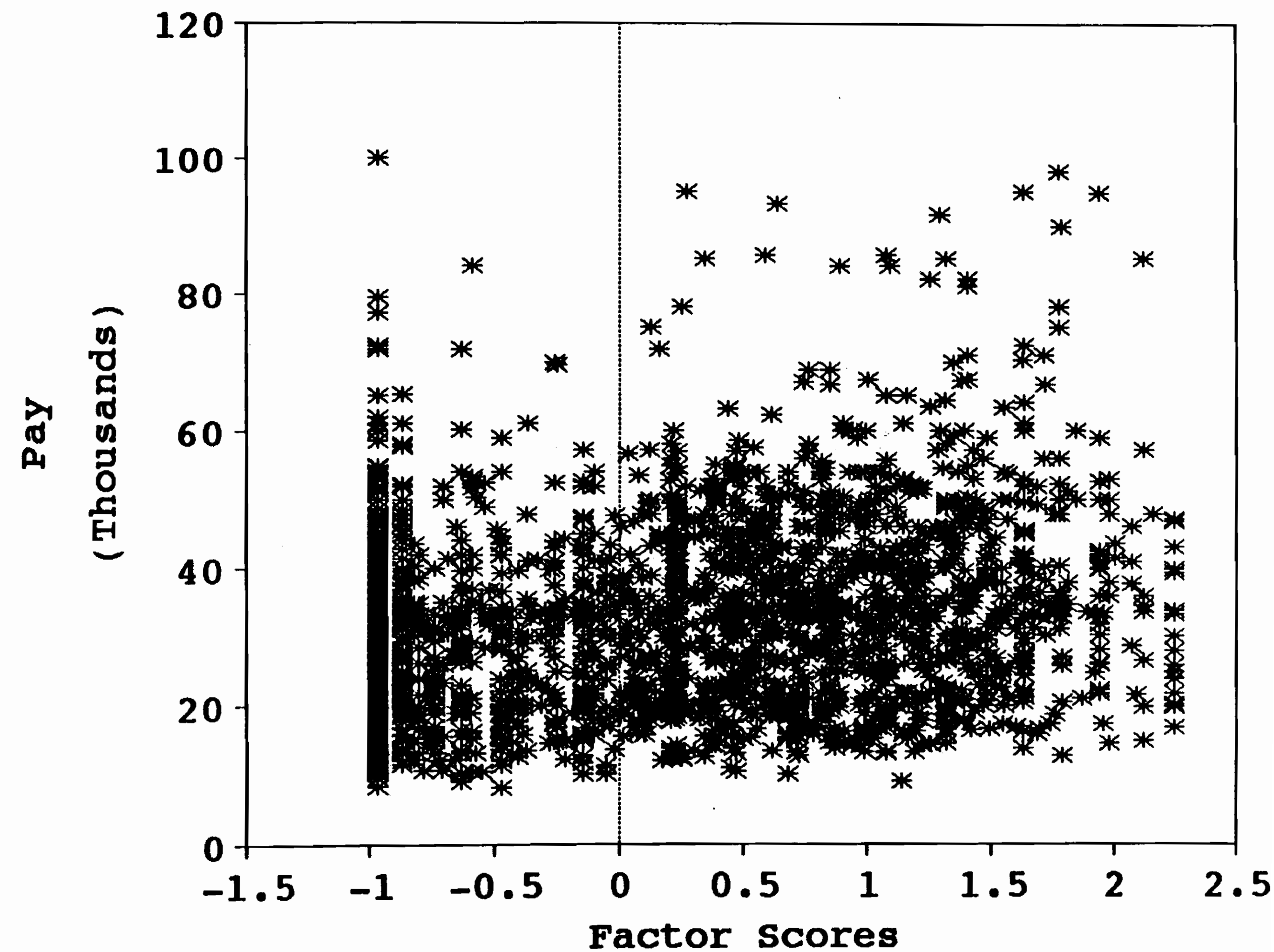
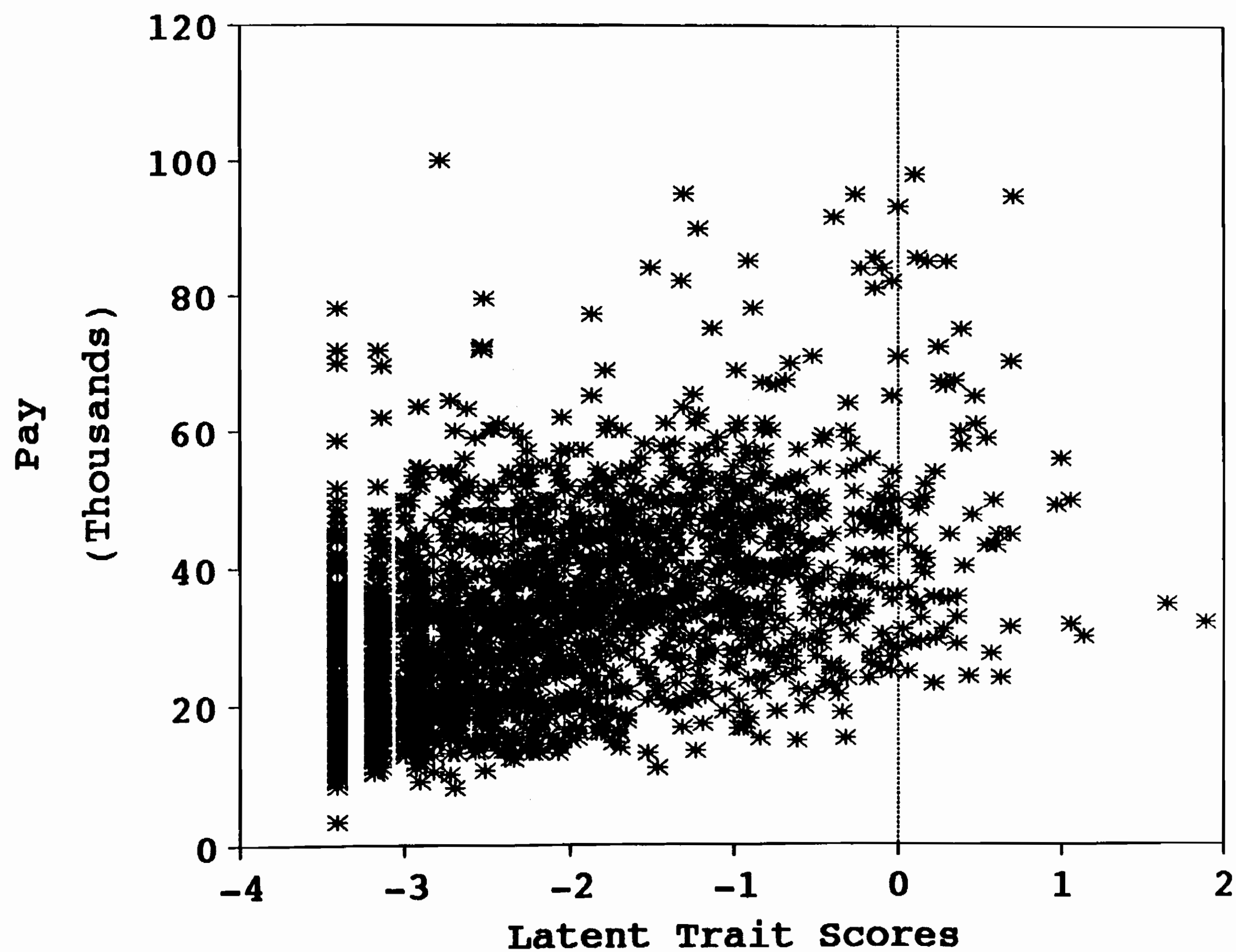
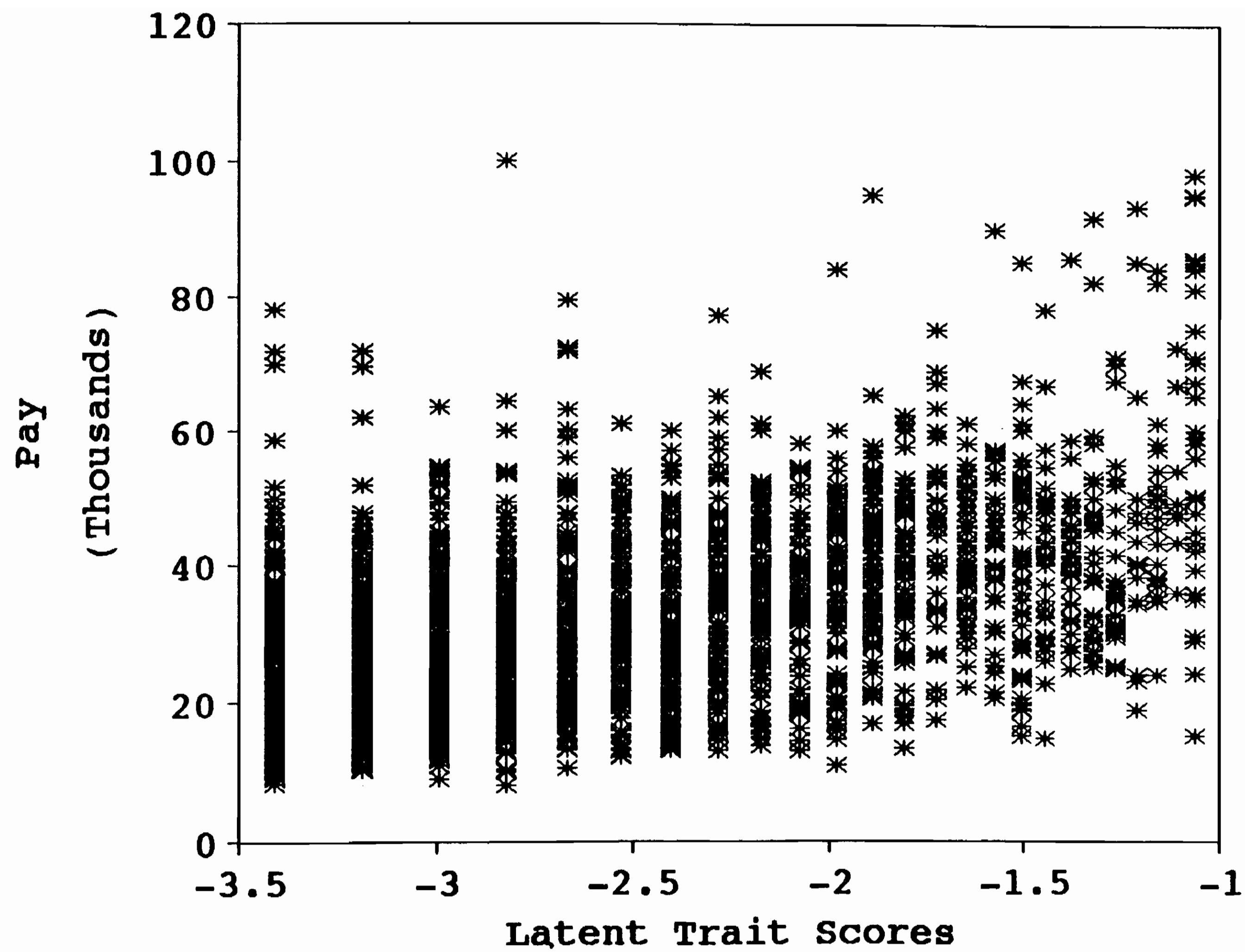


Figure 8. Scale 6 (factor scores by pay, frequency, dichotomous data).



<u>Figure 9.</u> Scale 2 (latent trait scores by pay, frequency, polychotomous data).



<u>Figure 10.</u> Scale 2 (latent trait scores by pay, frequency, dichotomous data).

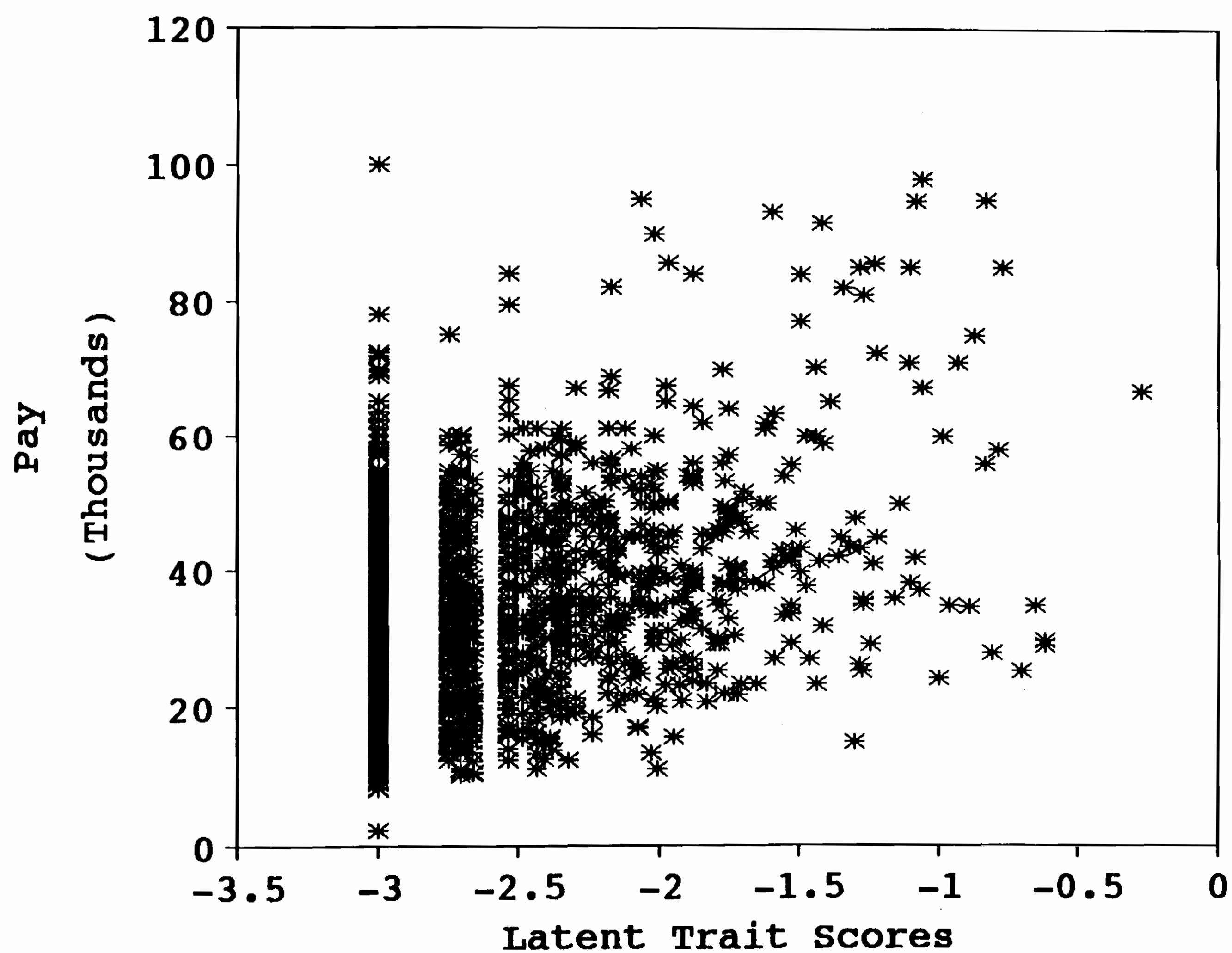


Figure 11. Scale 4 (latent trait scores by pay, frequency, polychotomous data).

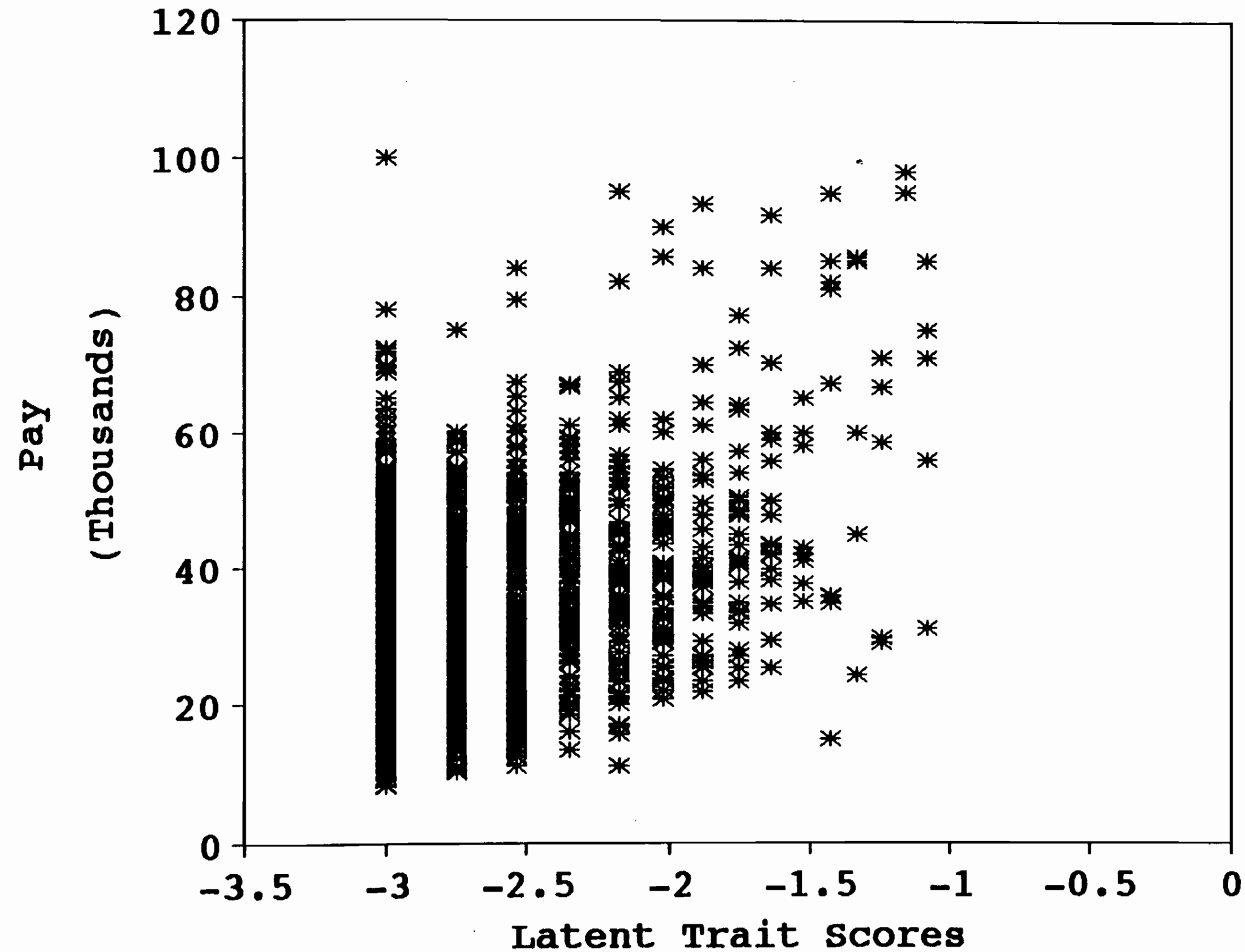


Figure 12. Scale 4 (latent trait scores by pay, frequency, dichotomous data).

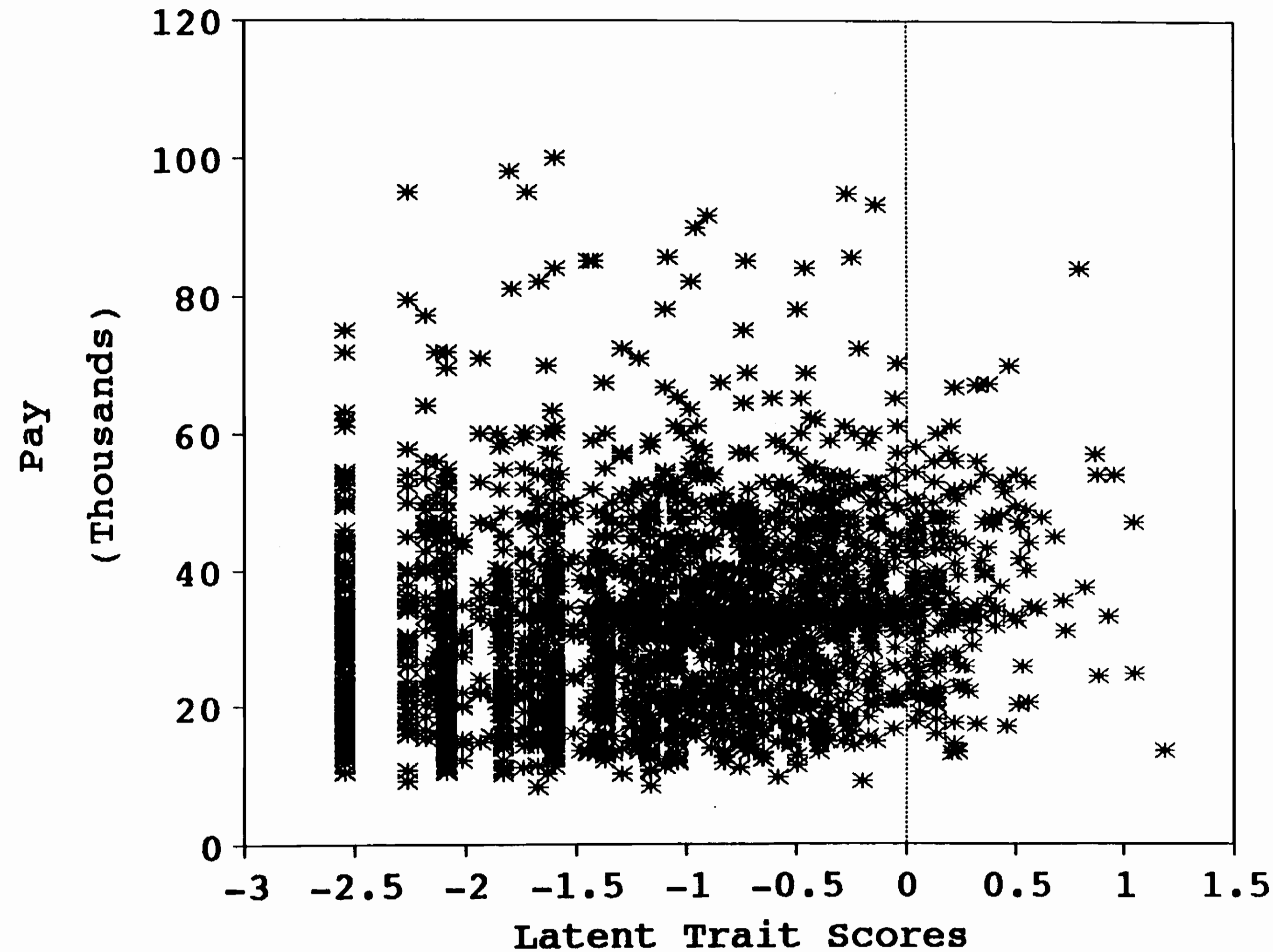


Figure 13. Scale 5 (latent trait scores by pay, frequency, polychotomous data).

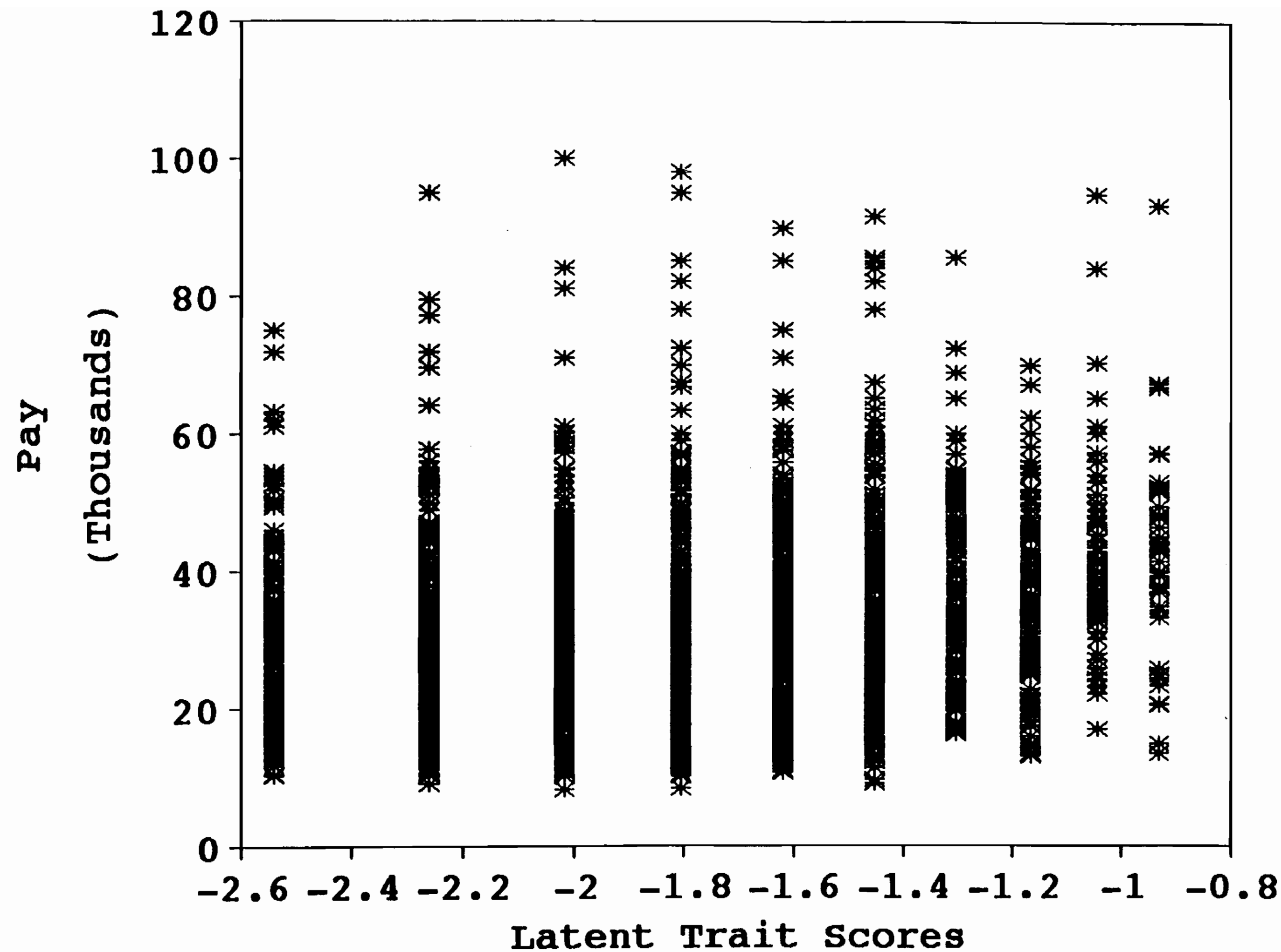


Figure 14. Scale 5 (latent trait scores by pay, frequency, dichotomous data).

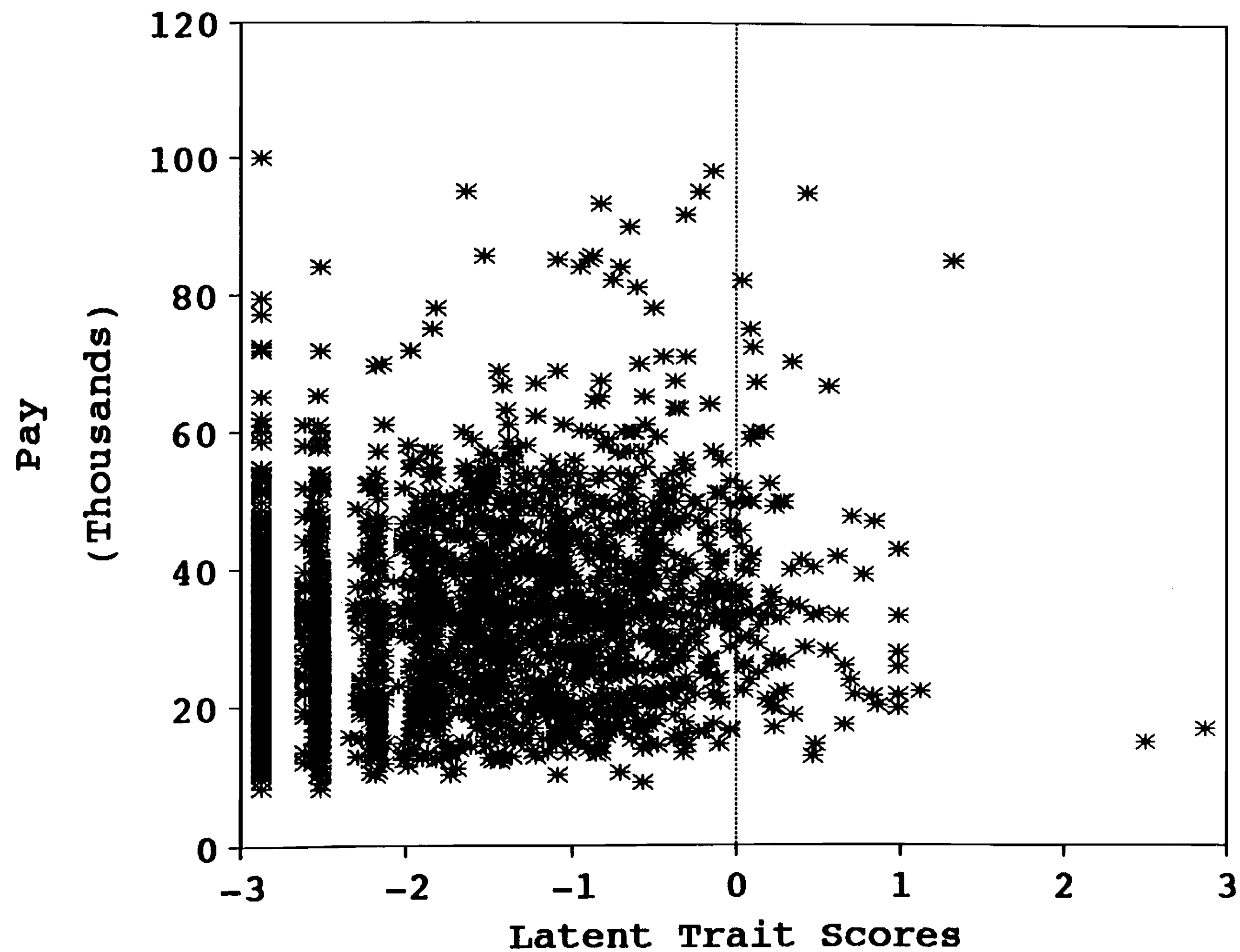


Figure 15. Scale 6 (latent trait scores by pay, frequency, polychotomous data).

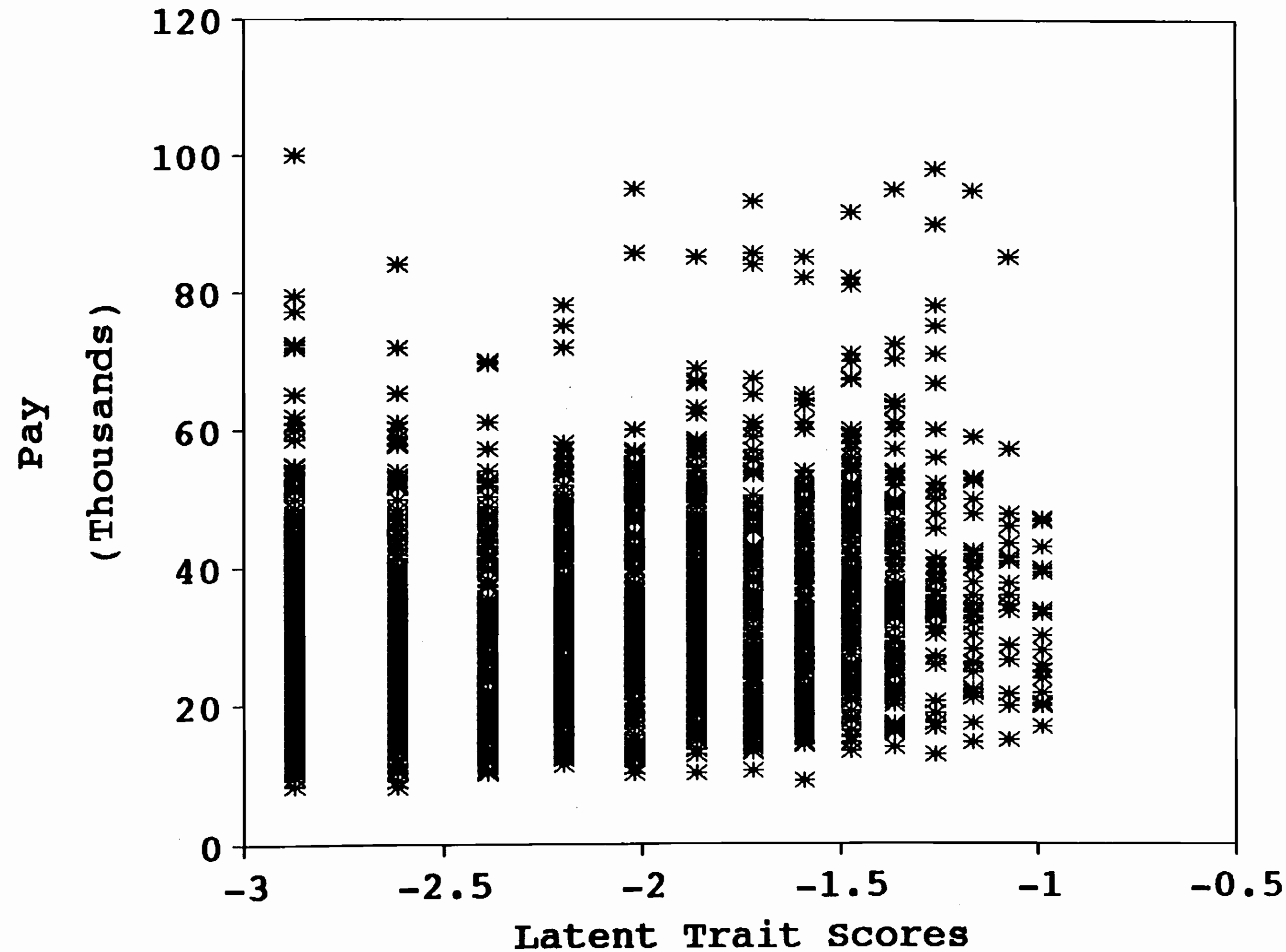


Figure 16. Scale 6 (latent trait scores by pay, frequency, dichotomous data).

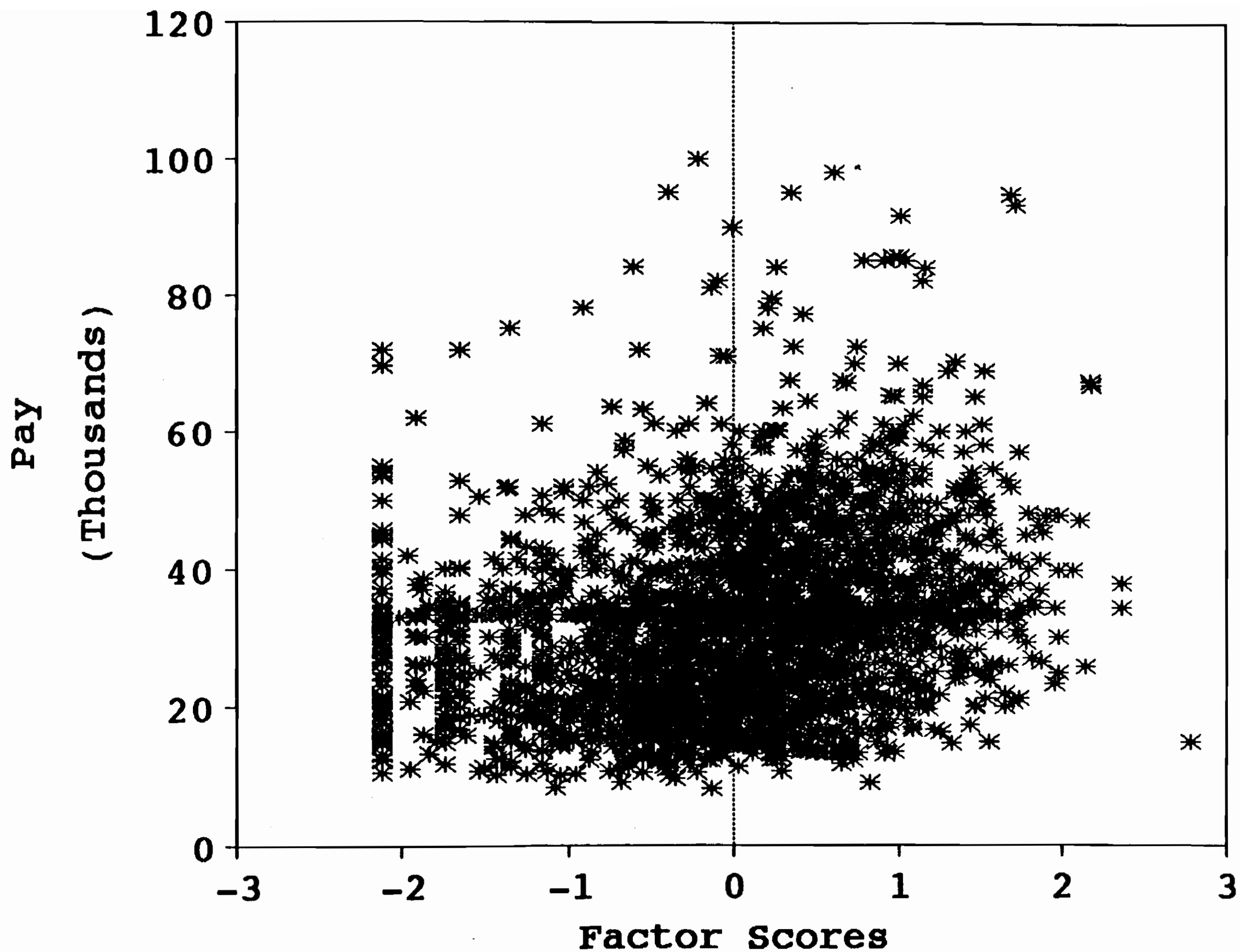


Figure 17. Scale 7 (factor scores by pay, frequency, polychotomous data).

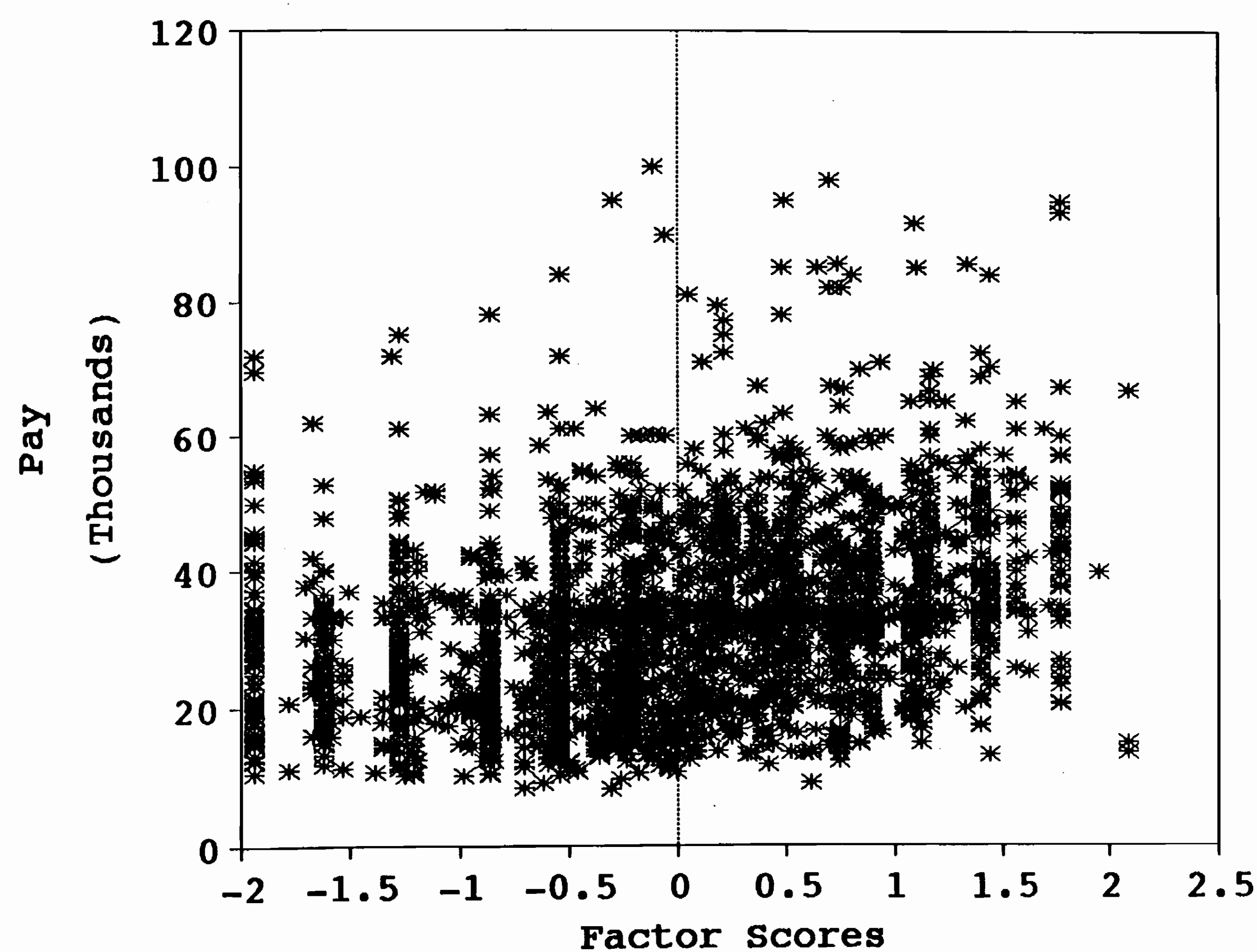


Figure 18. Scale 7 (factor scores by pay, frequency, dichotomous data).

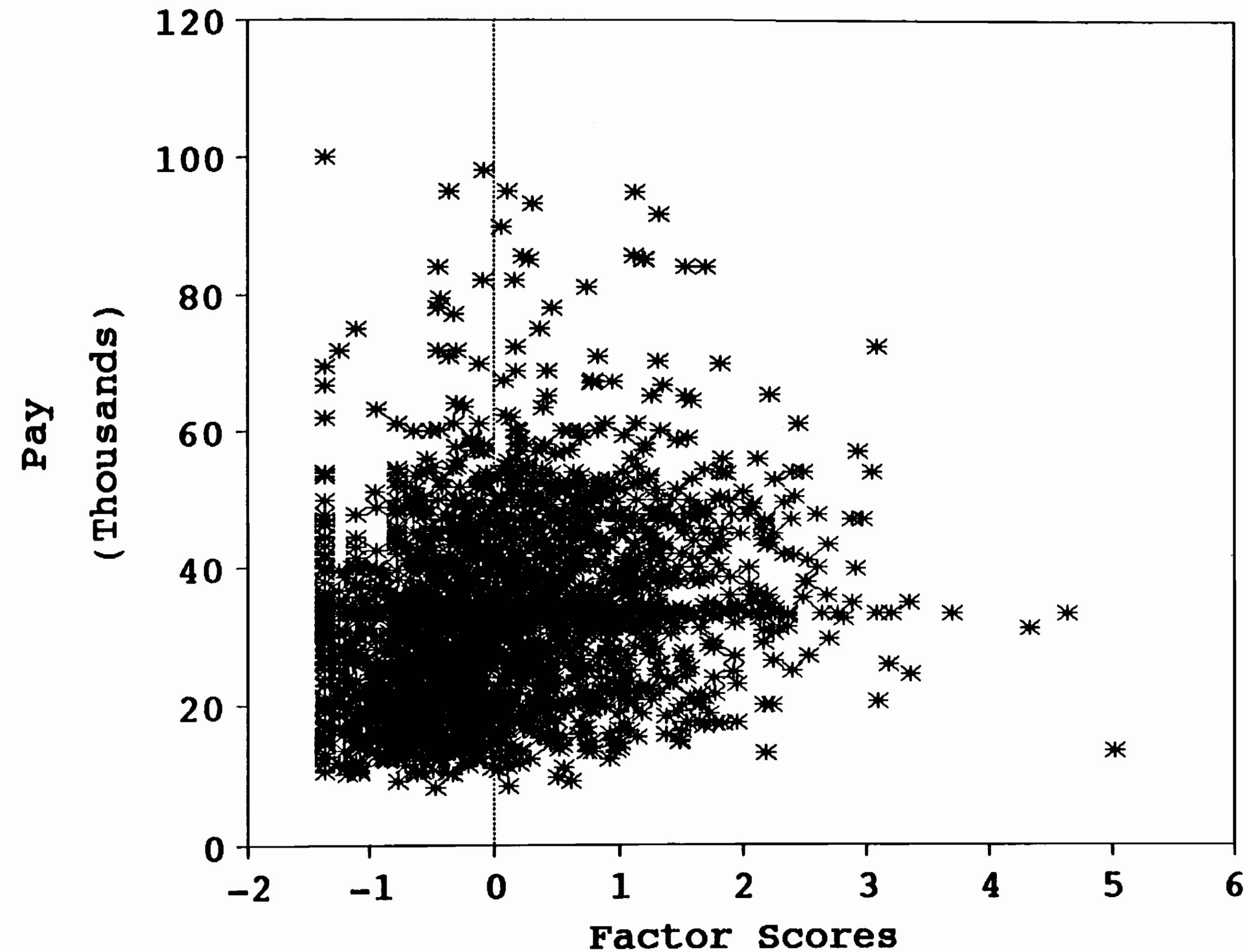
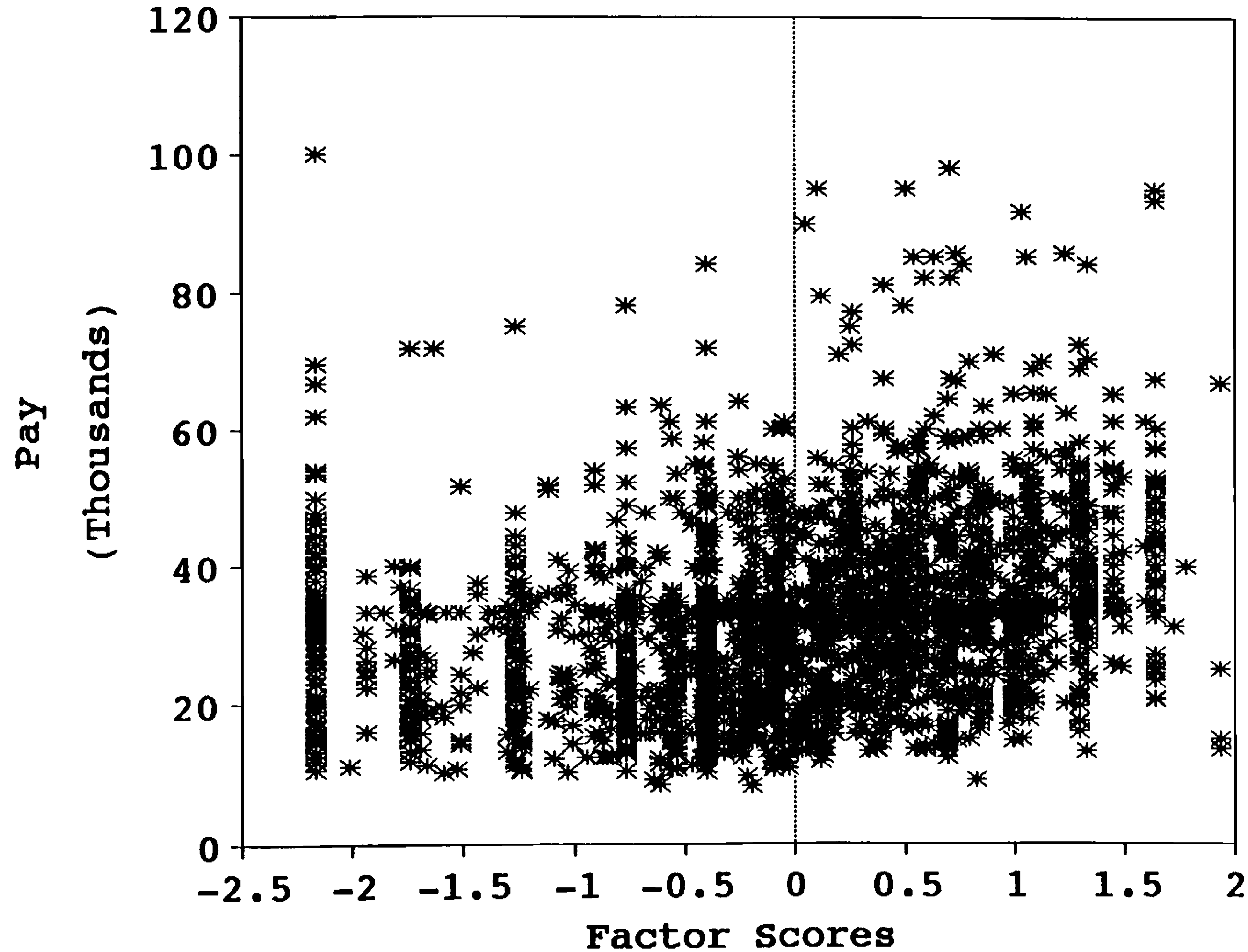


Figure 19. Scale 7 (factor scores by pay, importance, polychotomous data).



<u>Figure 20.</u> Scale 7 (factor scores by pay, importance, dichotomous data).

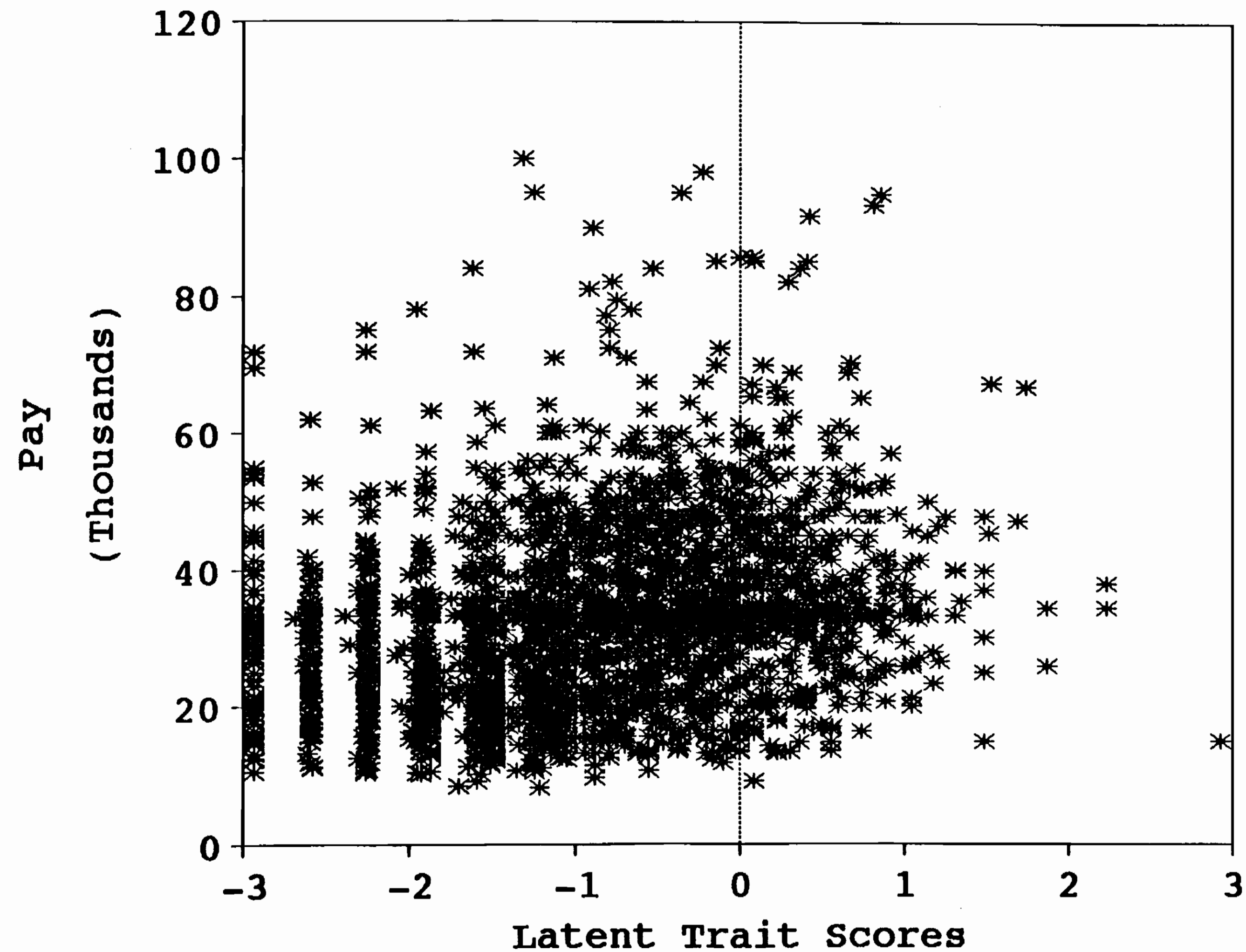


Figure 21. Scale 7 (latent trait scores by pay, frequency, polychotomous data).

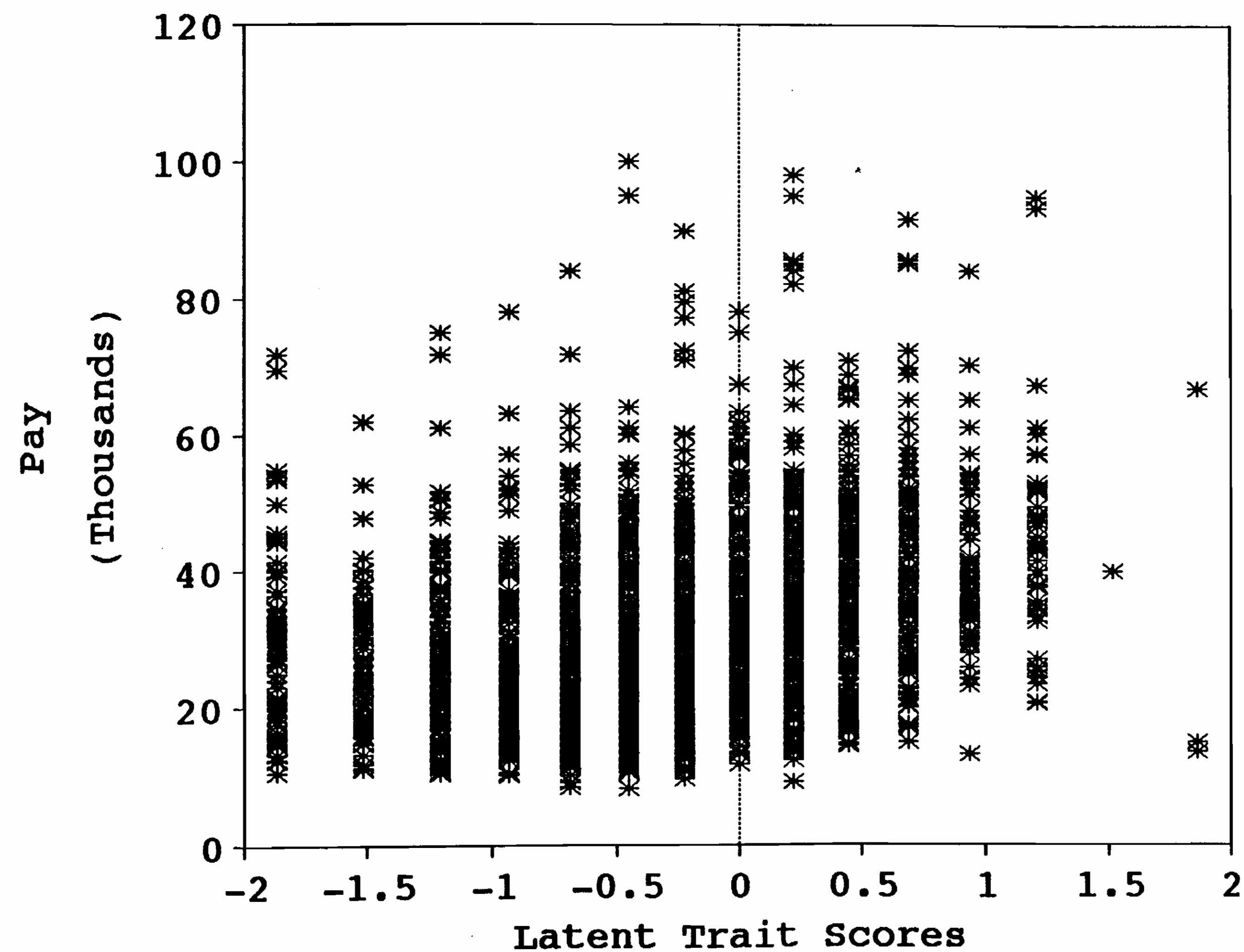
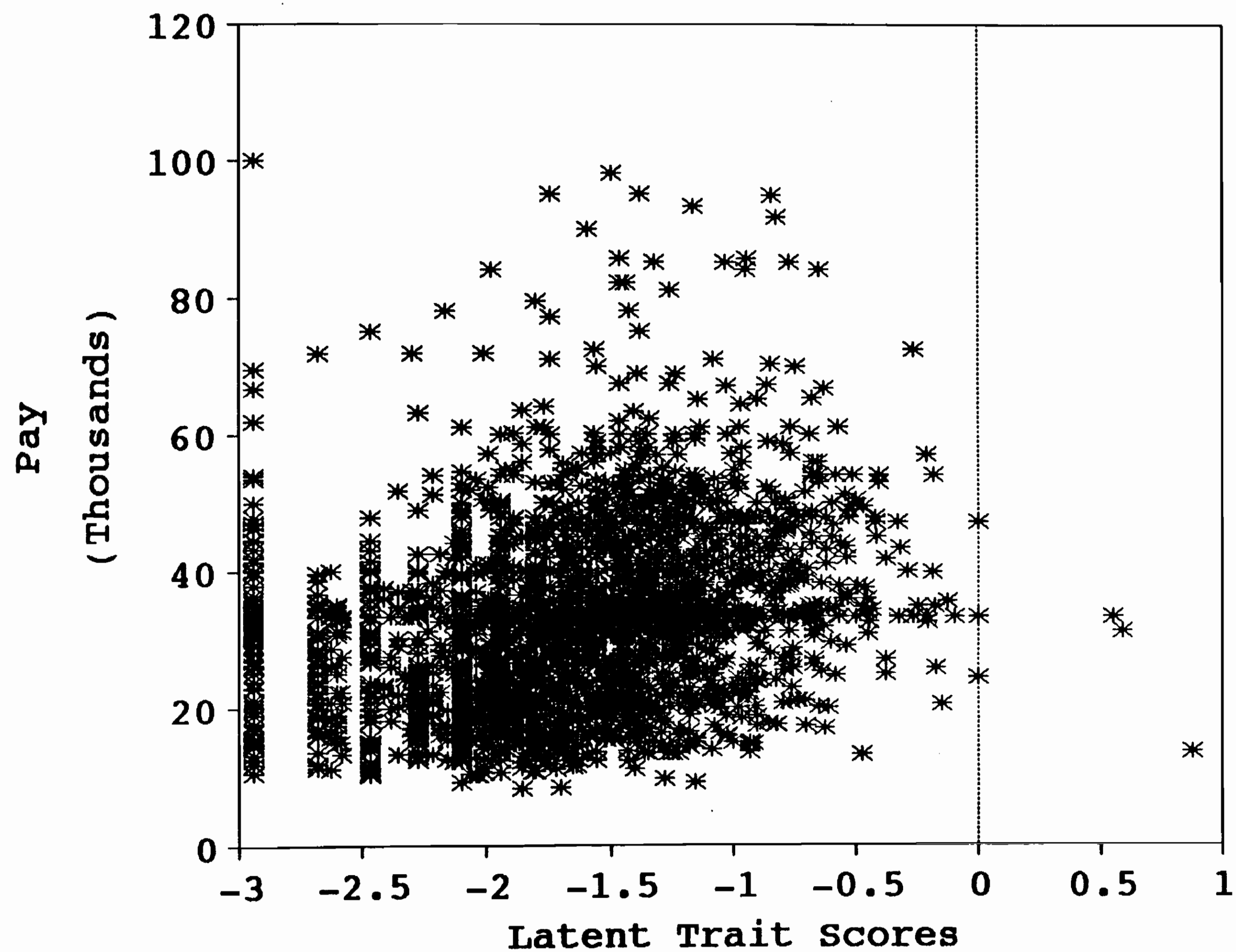


Figure 22. Scale 7 (latent trait scores by pay, frequency, dichotomous data).



<u>Figure 23.</u> Scale 7 (latent trait scores by pay, importance, polychotomous data).

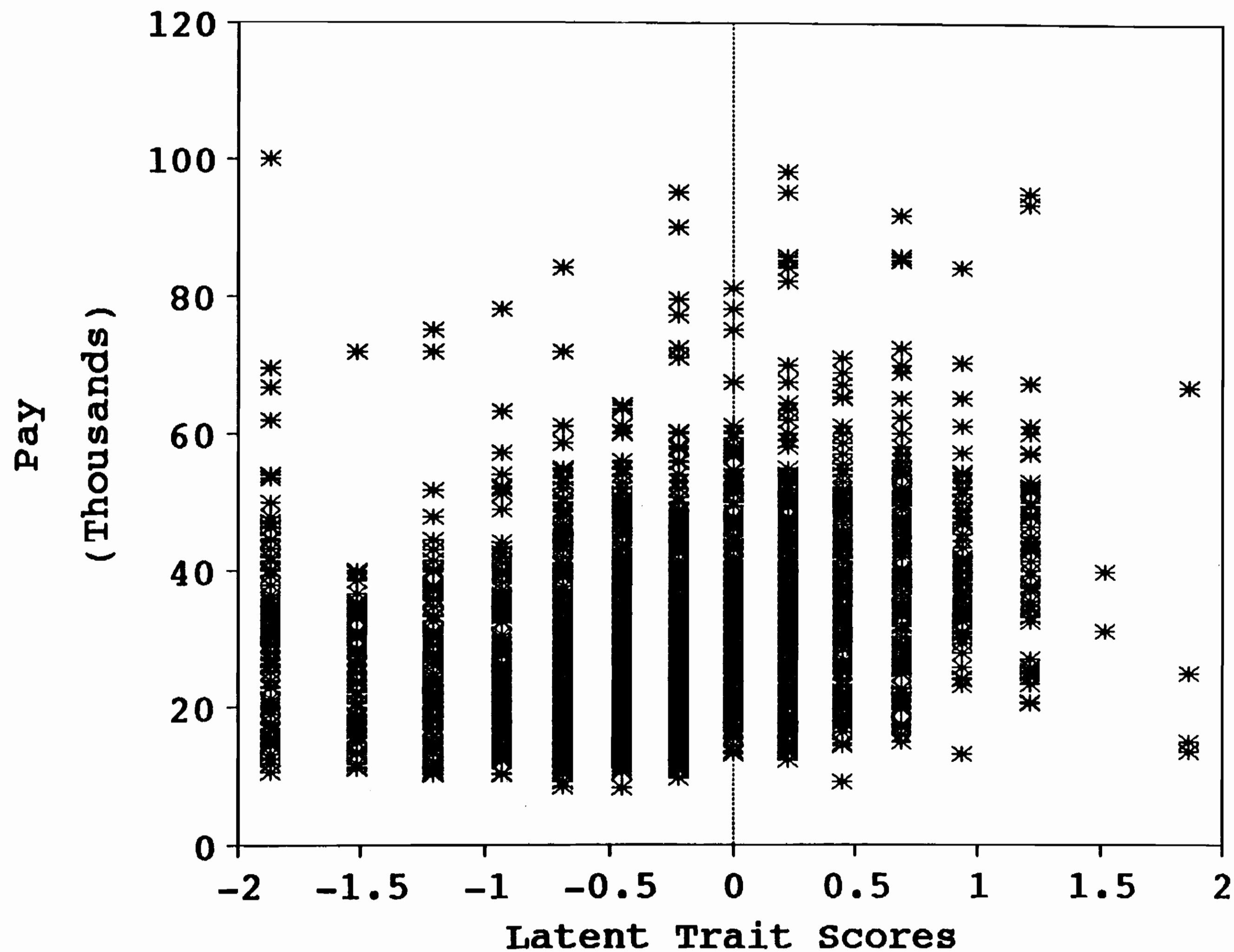


Figure 24. Scale 7 (latent trait scores by pay, importance, dichotomous data).

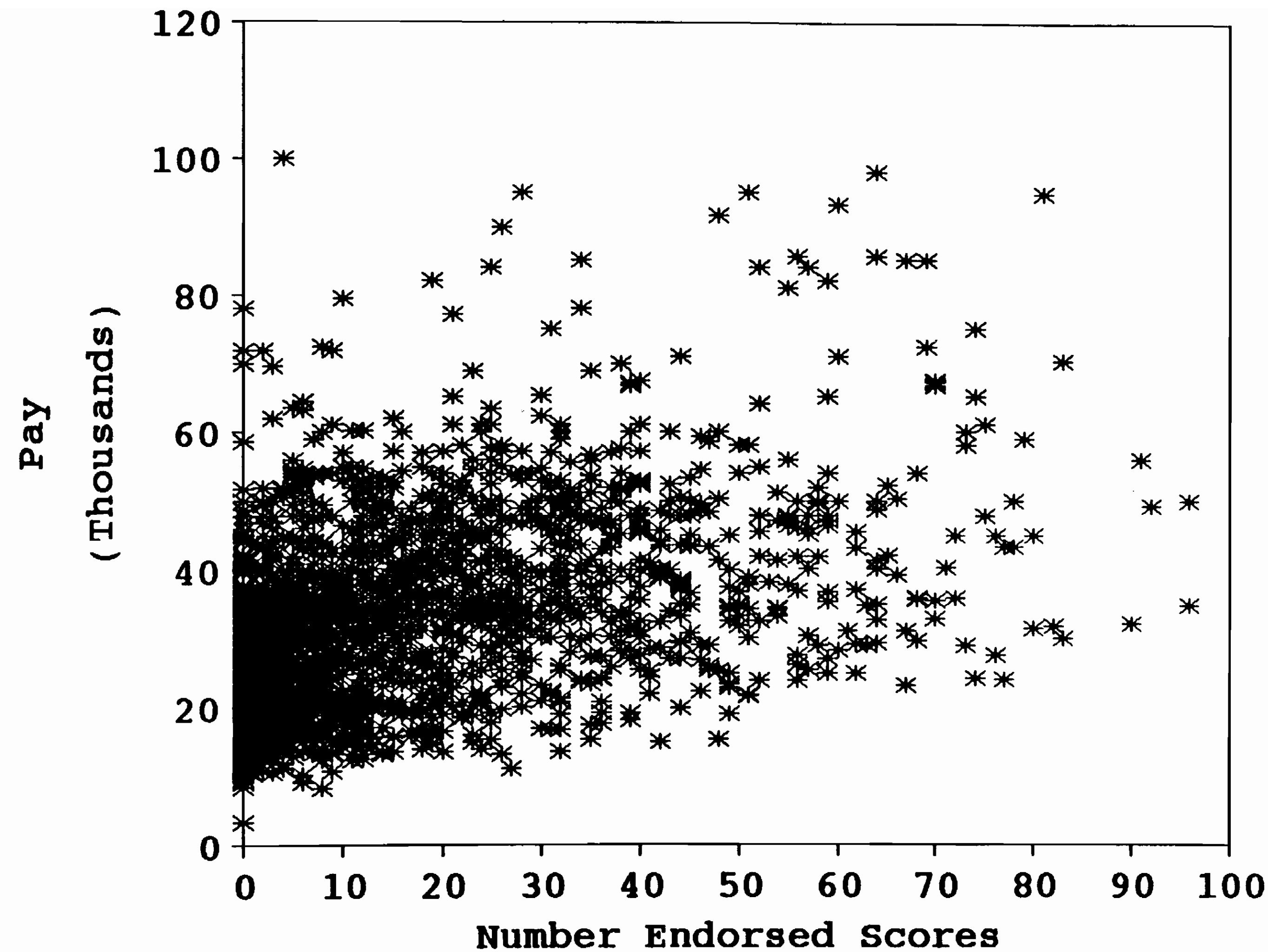


Figure 25. Scale 2 (number endorsed scores by pay, polychotomous data).

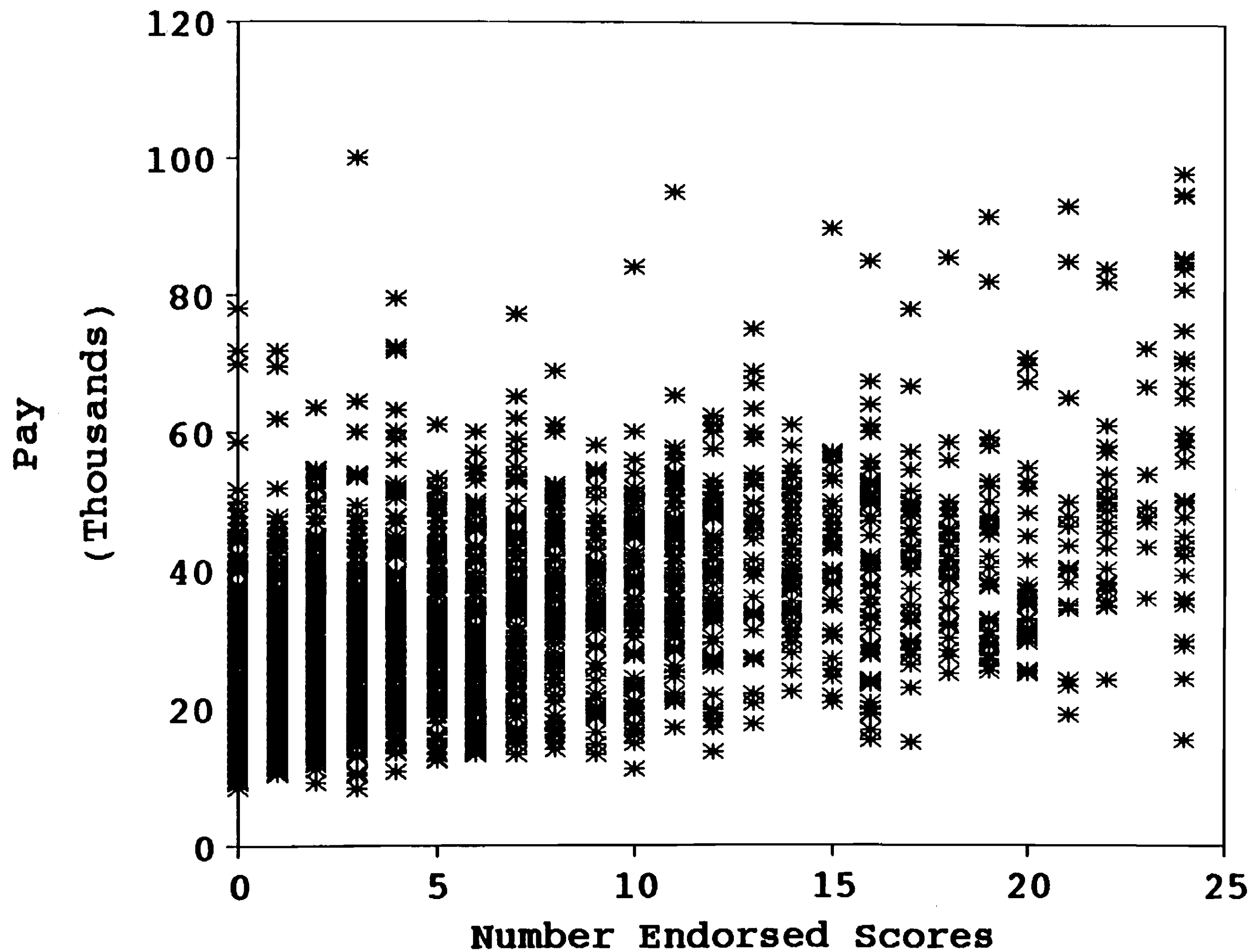


Figure 26. Scale 2 (number endorsed scores by pay, dichotomous data).

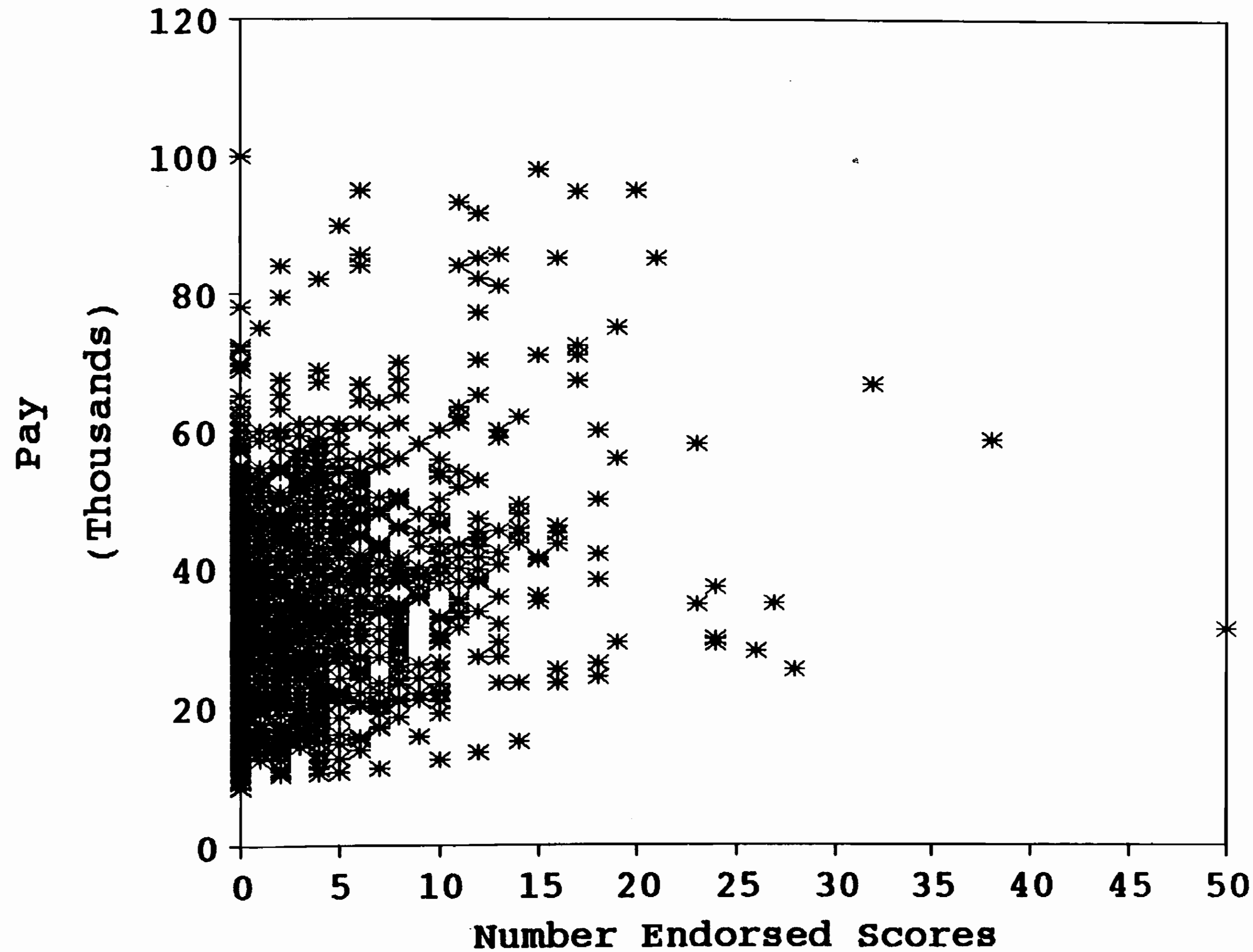


Figure 27. Scale 4 (number endorsed scores by pay, polychotomous data).

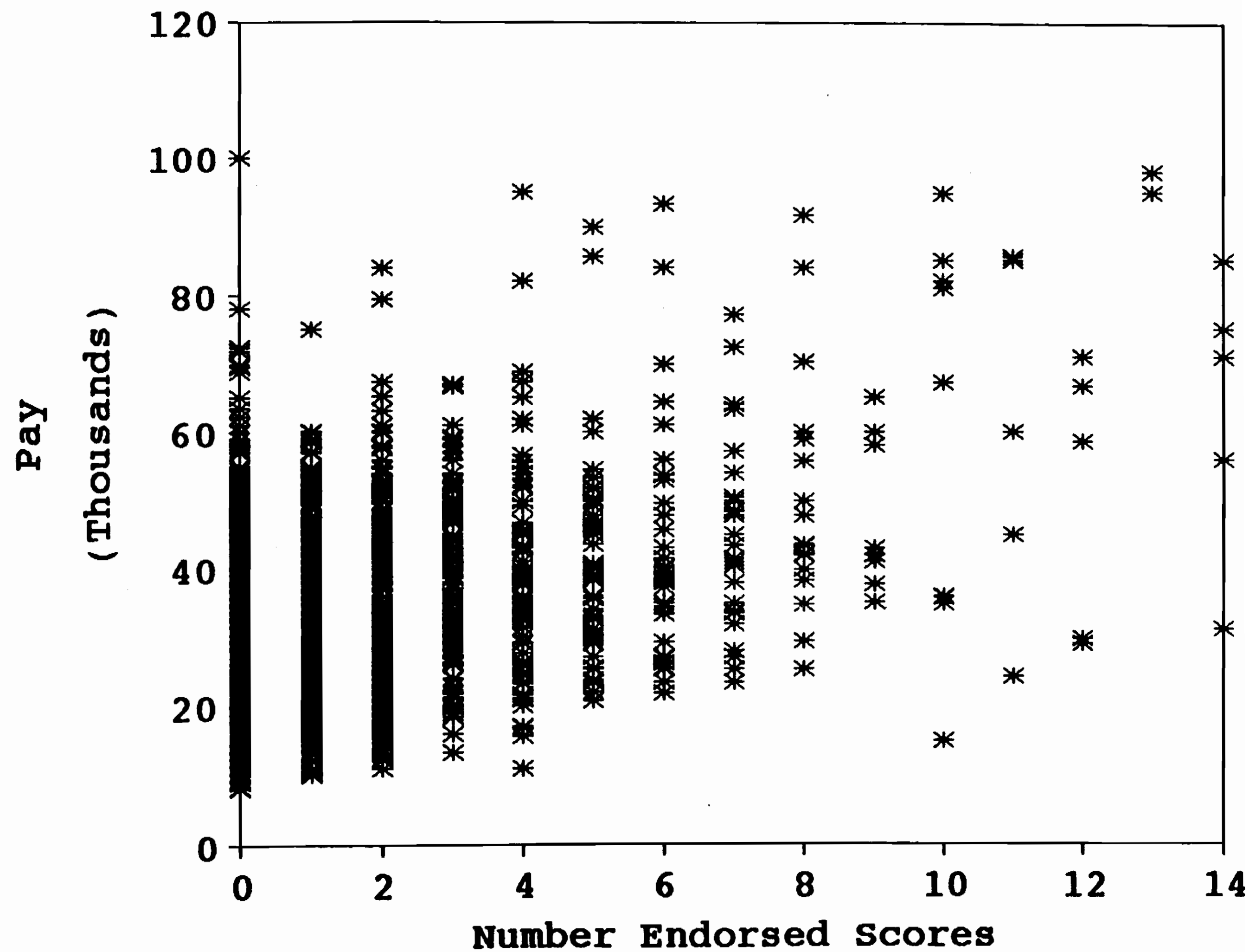


Figure 28. Scale 4 (number endorsed scores by pay, dichotomous data).

£4

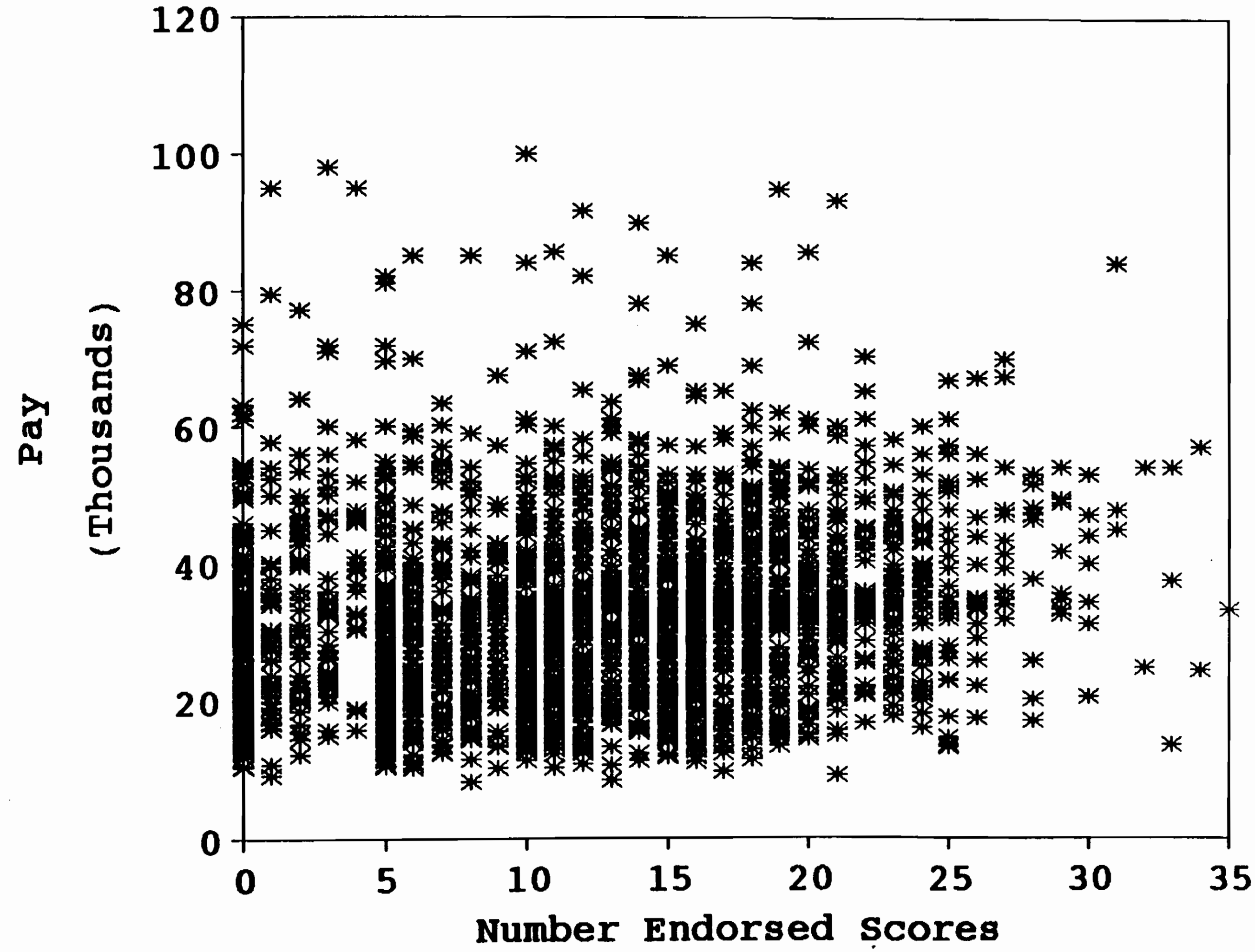
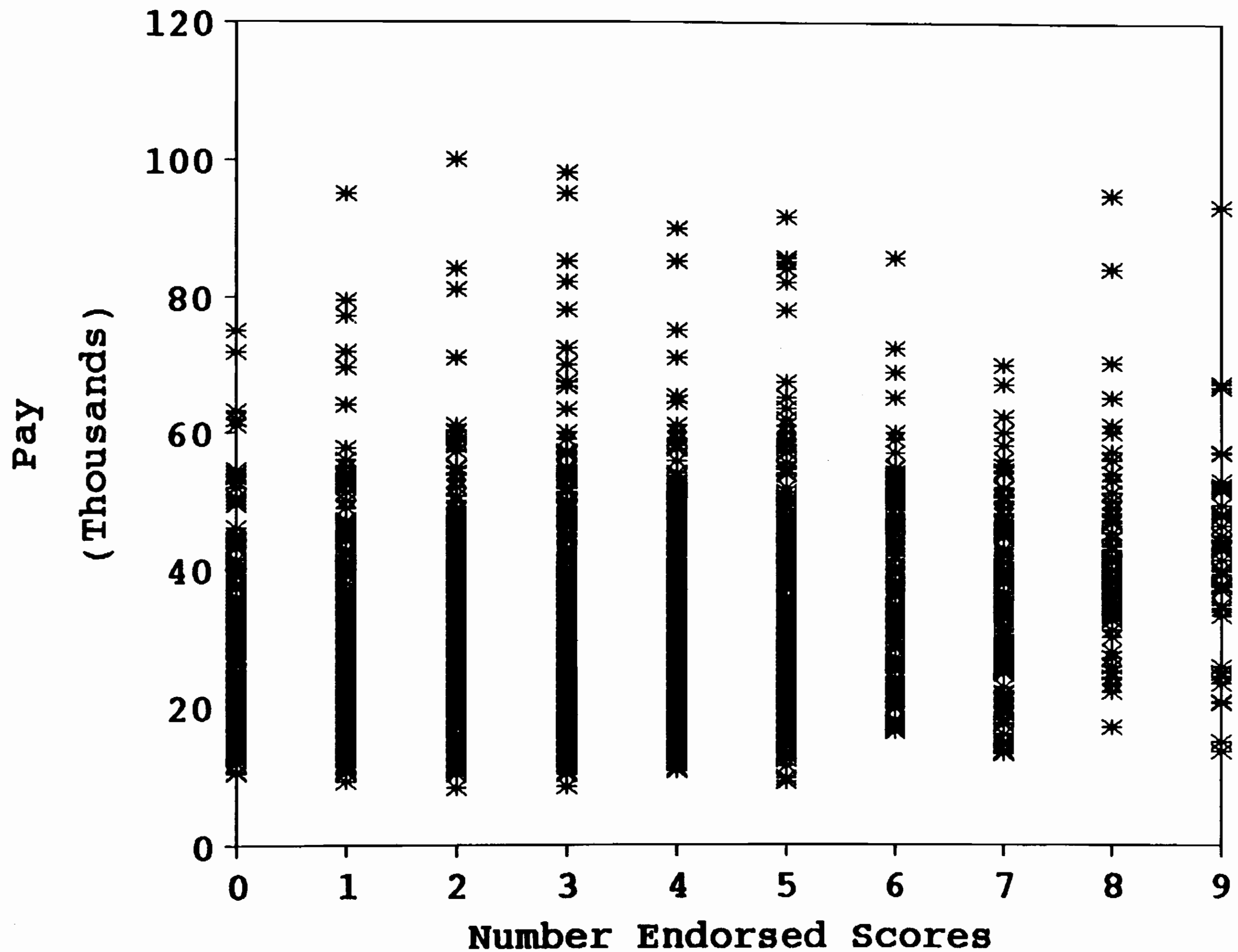
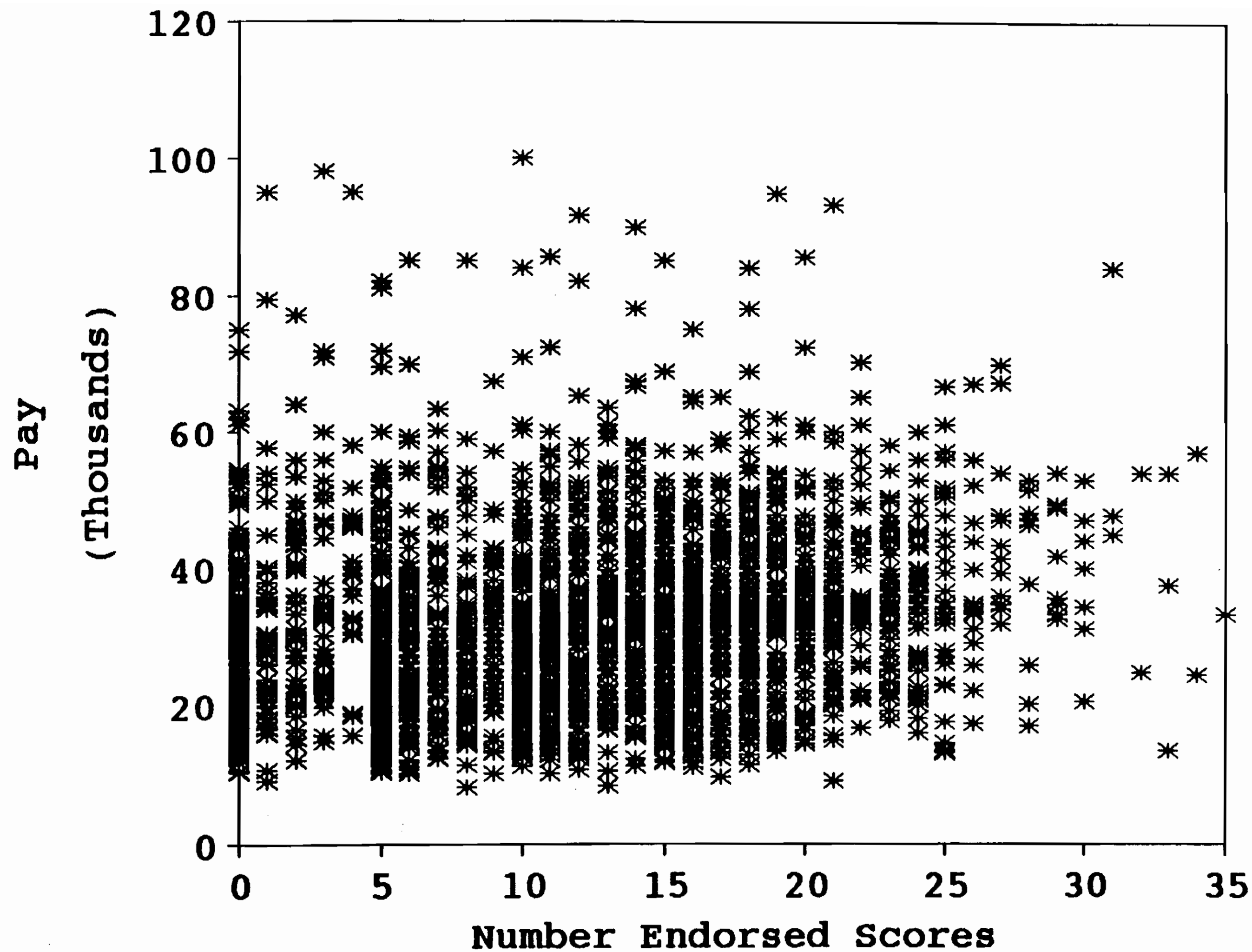


Figure 29. Scale 5 (number endorsed scores by pay, polychotomous data).



<u>Figure 30.</u> Scale 5 (number endorsed scores by pay, dichotomous data).



<u>Figure 31.</u> Scale 6 (number endorsed scores by pay, polychotomous data).

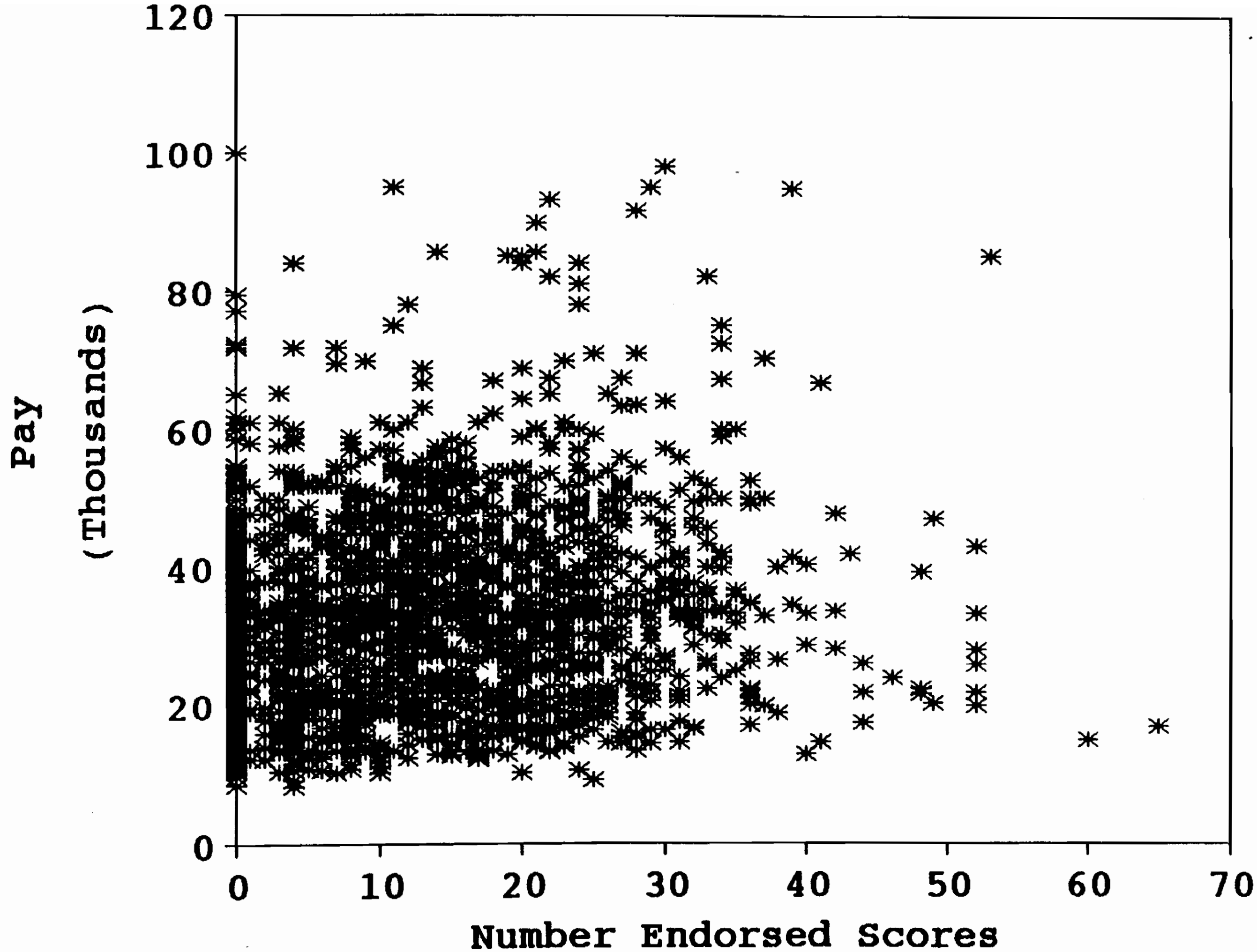


Figure 32. Scale 6 (number endorsed scores by pay, dichotomous data).

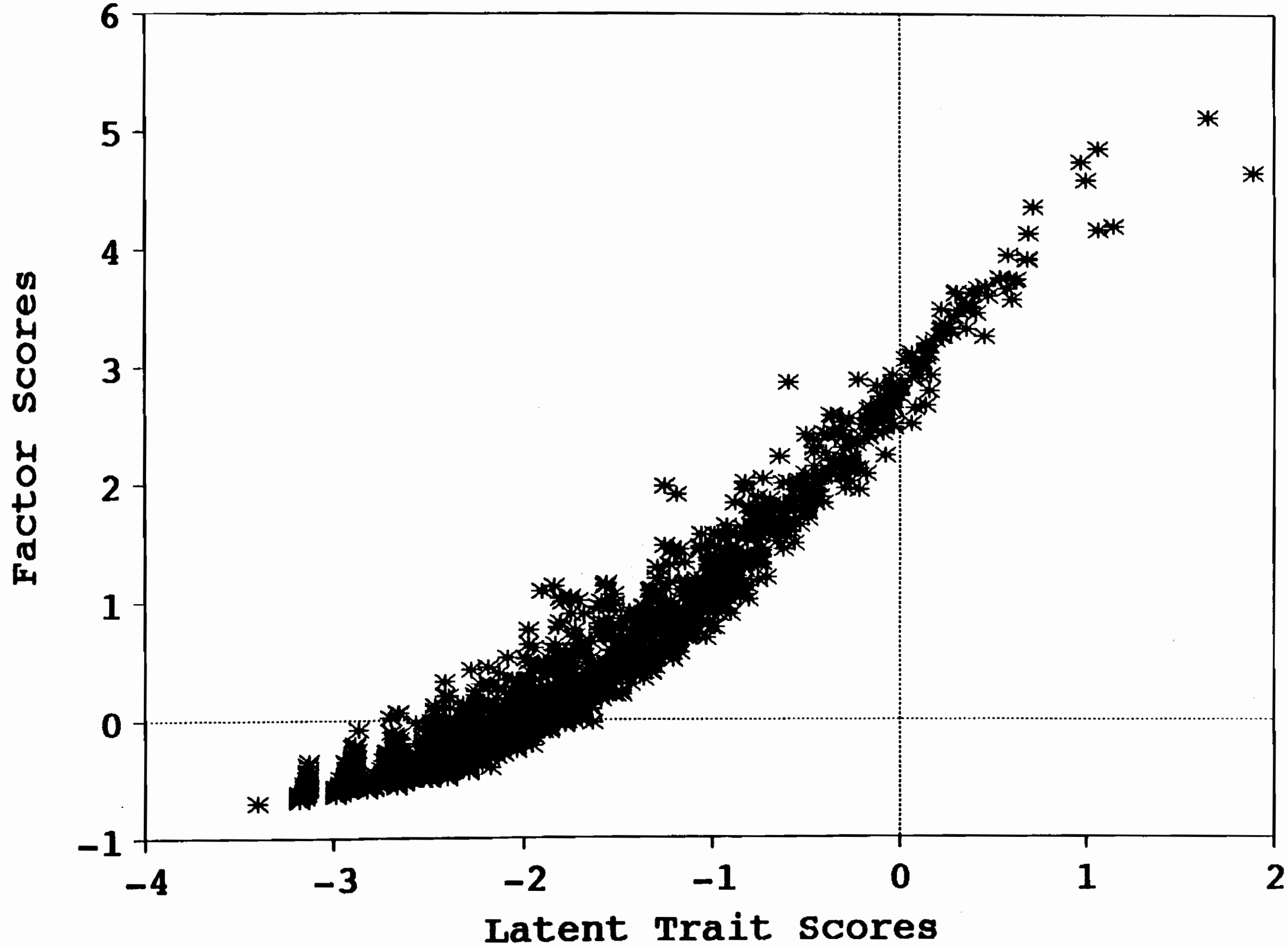
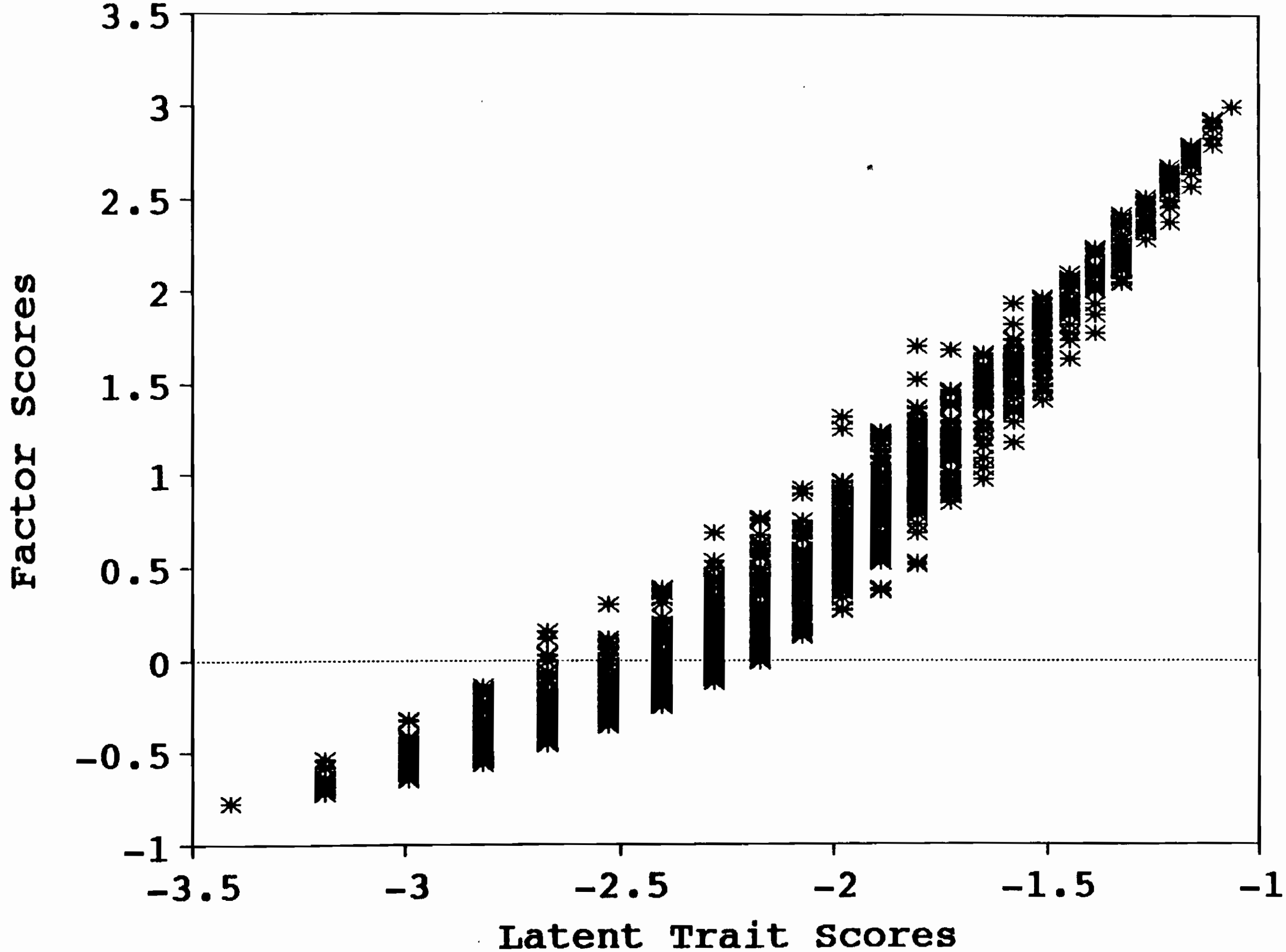


Figure 33. Scale 2 (latent trait by factor scores, L-2, polychotomous data).



<u>Figure 34.</u> Scale 2 (latent trait by factor scores, L-2, dichotomous data).

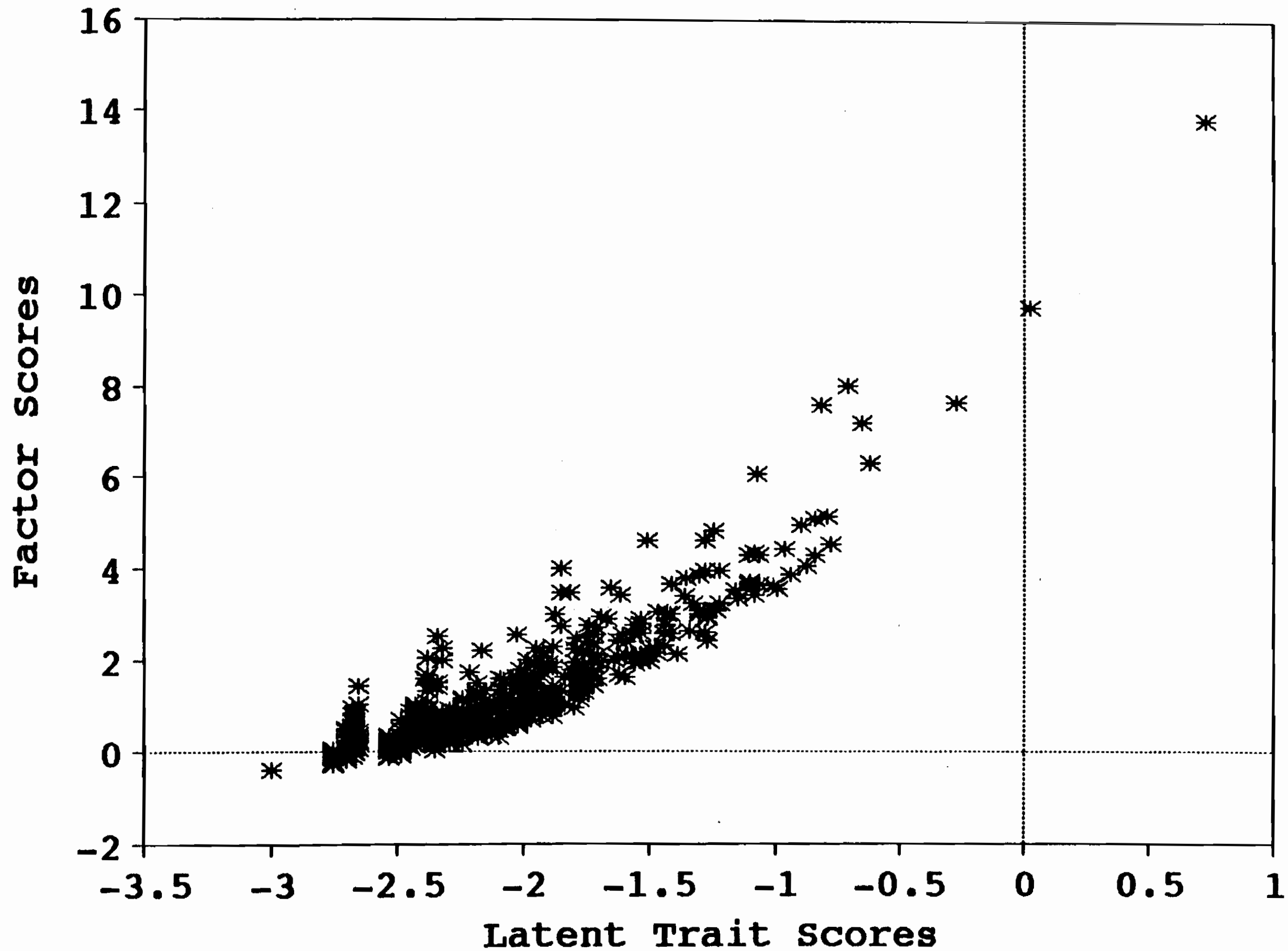
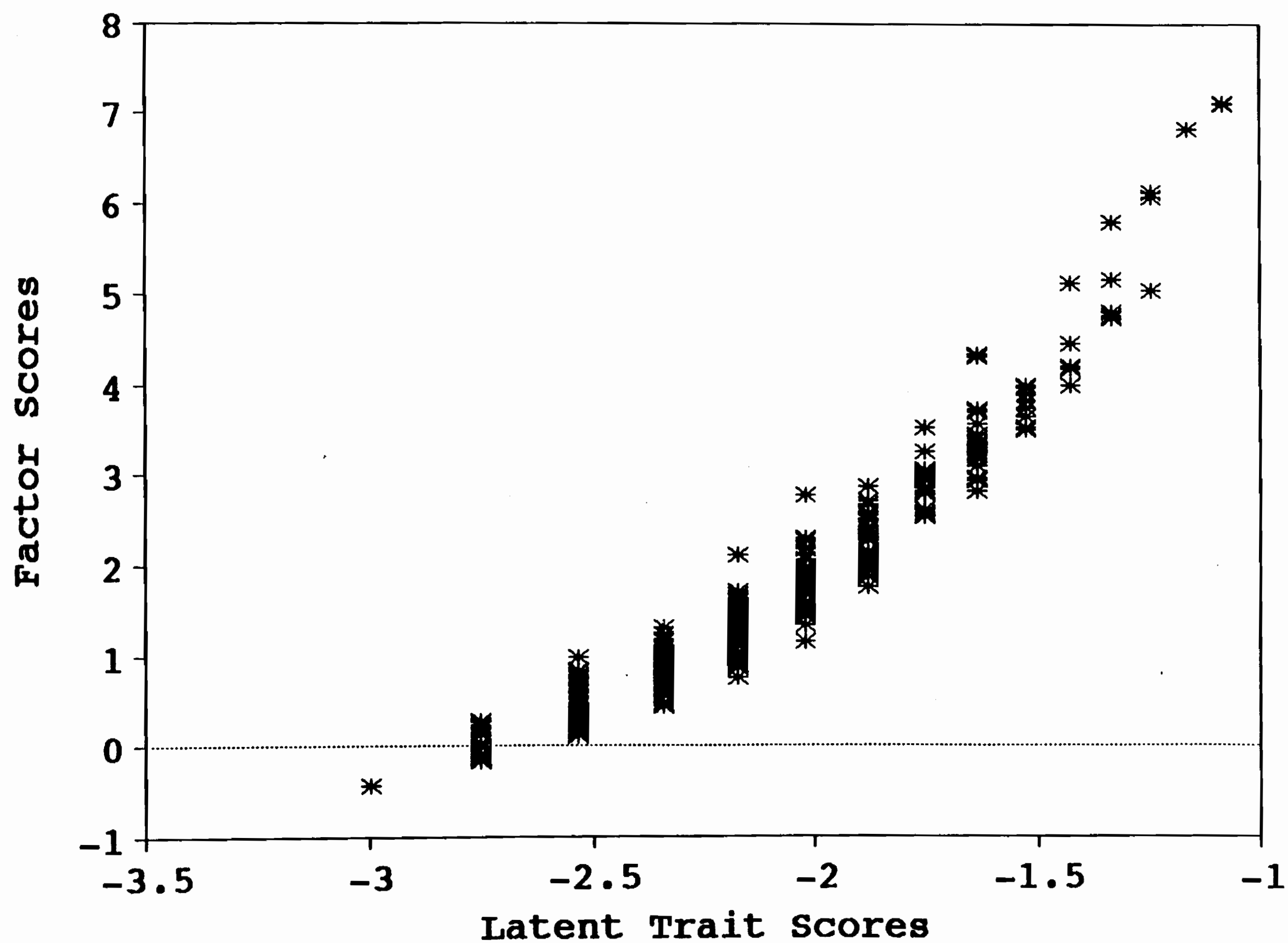


Figure 35. Scale 4 (latent trait by factor scores, L-2, polychotomous data).



<u>Figure 36.</u> Scale 4 (latent trait by factor scores, L-2, dichotomous data).

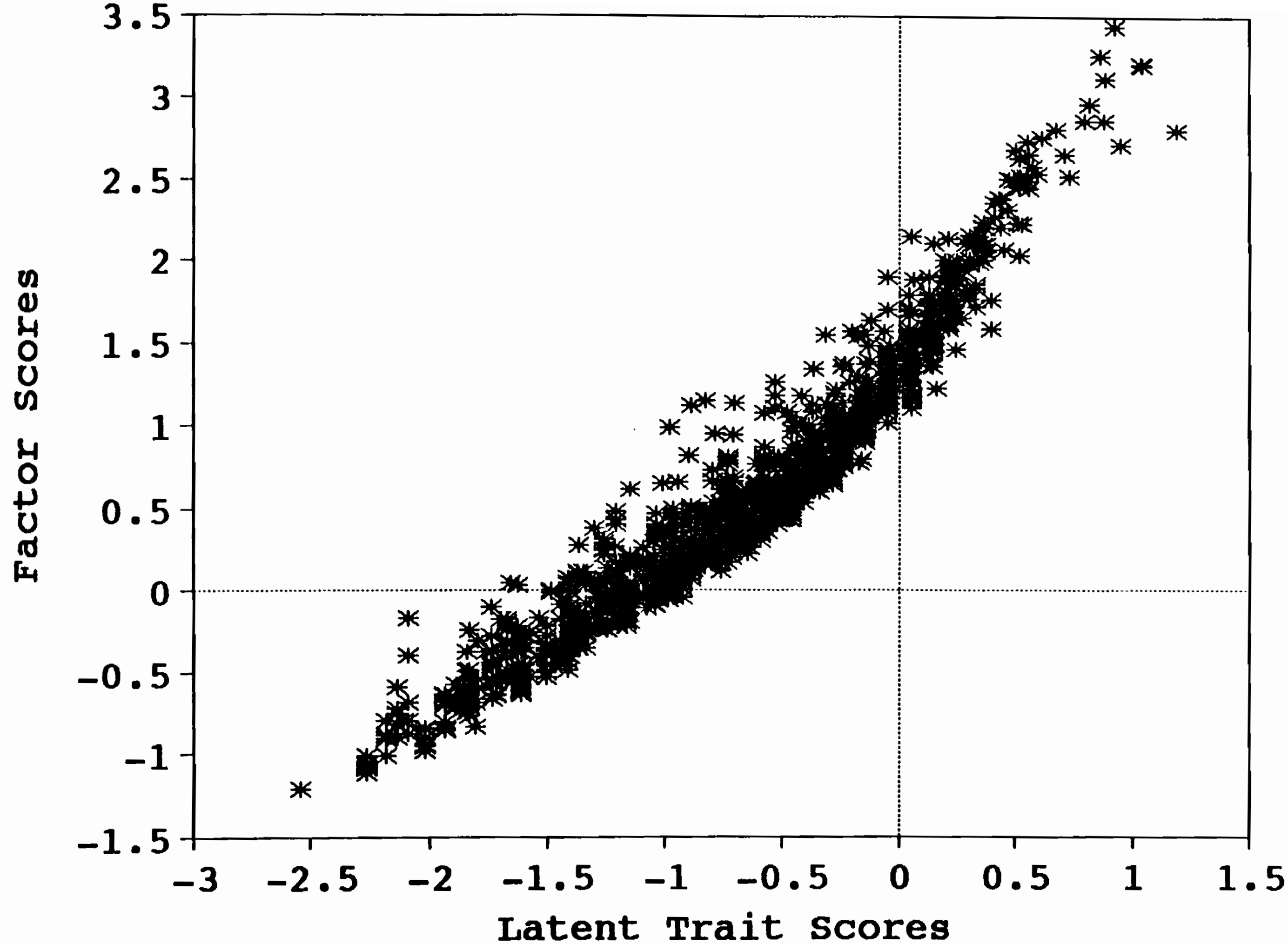
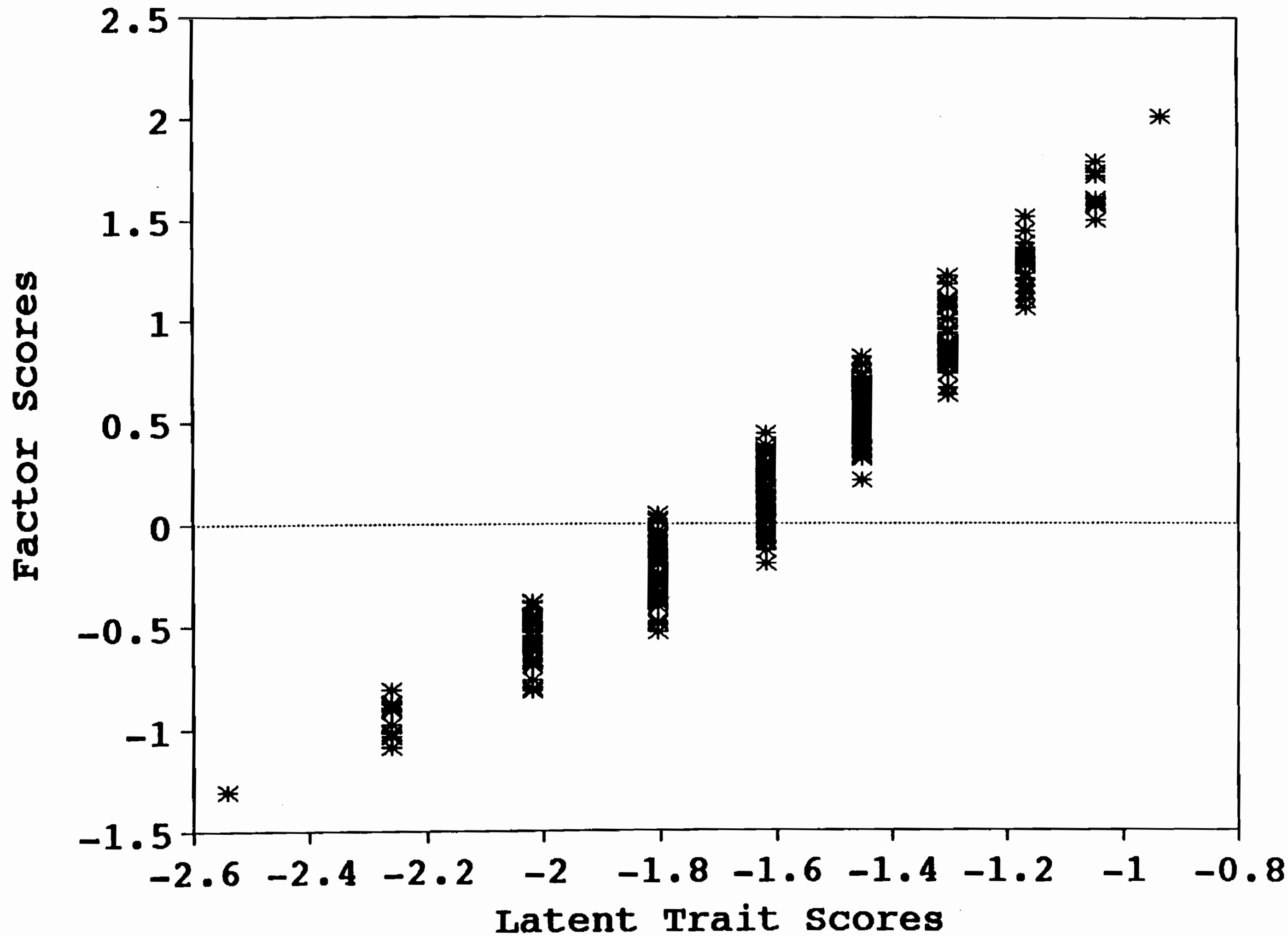


Figure 37. Scale 5 (latent trait by factor scores, L-2, polychotomous data).



<u>Figure 38.</u> Scale 5 (latent trait by factor scores, L-2, dichotomous data).

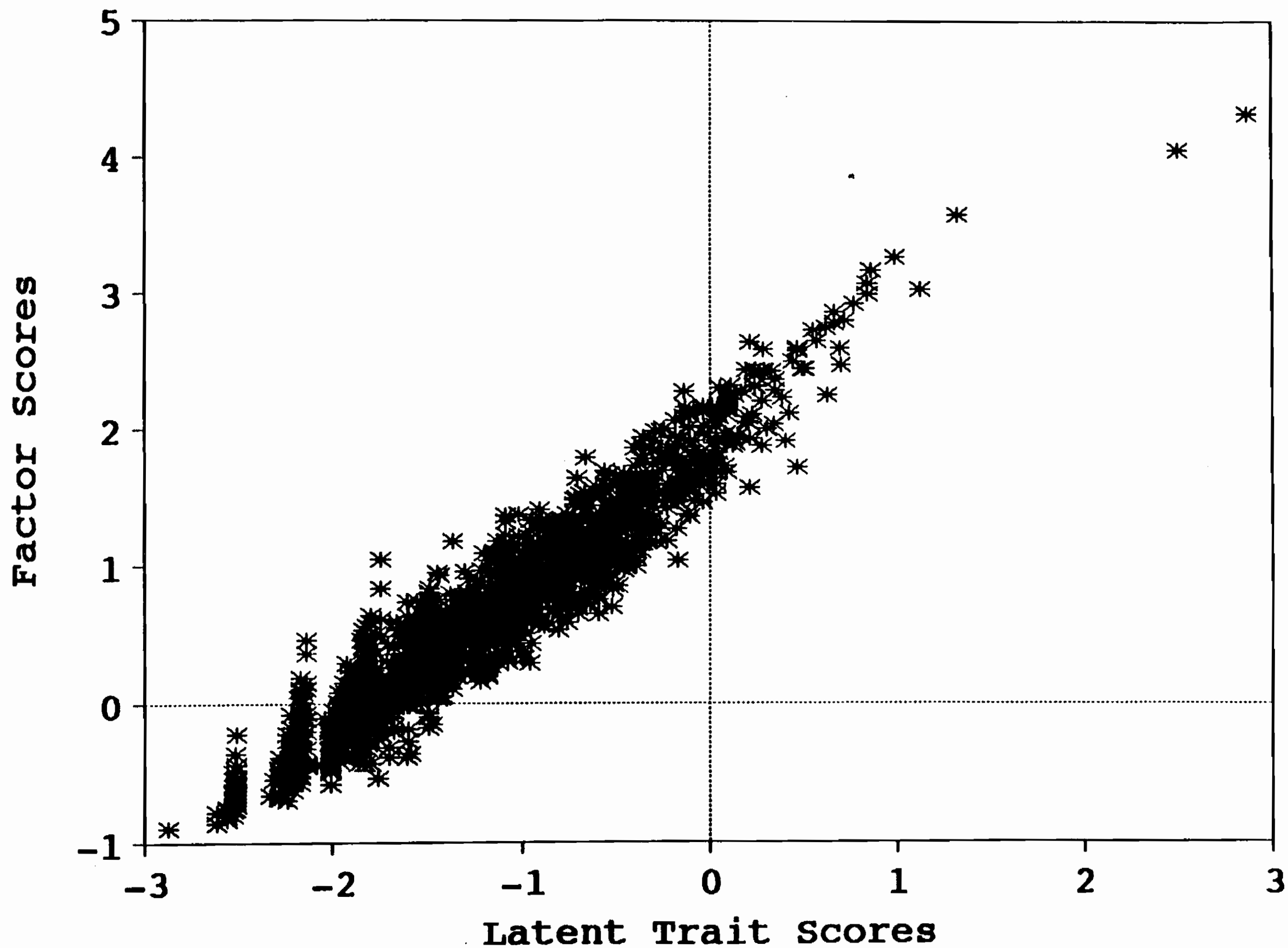


Figure 39. Scale 6 (latent trait by factor scores, L-2, polychotomous data).

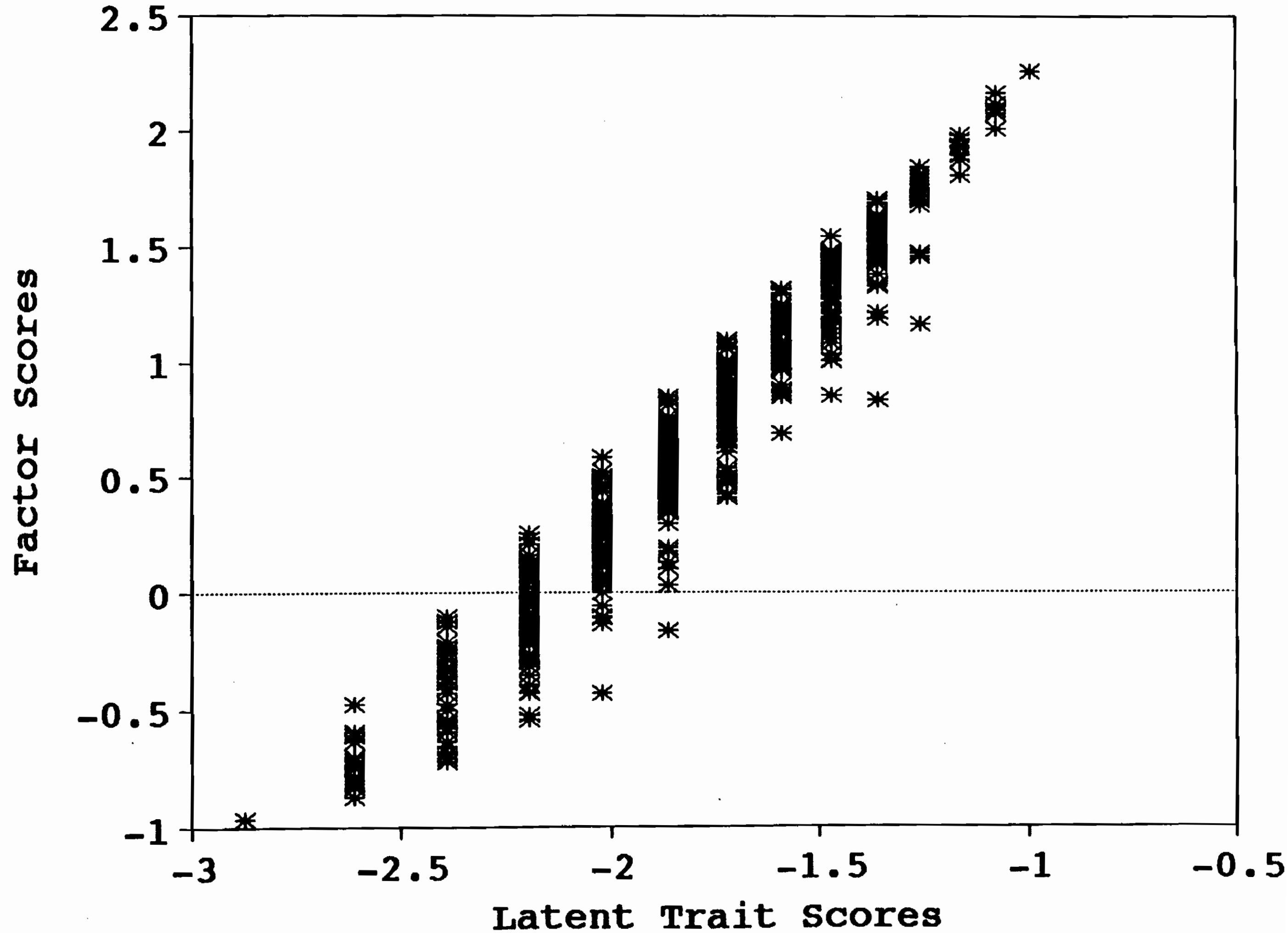


Figure 40. Scale 6 (latent trait by factor scores, L-2, dichotomous data).

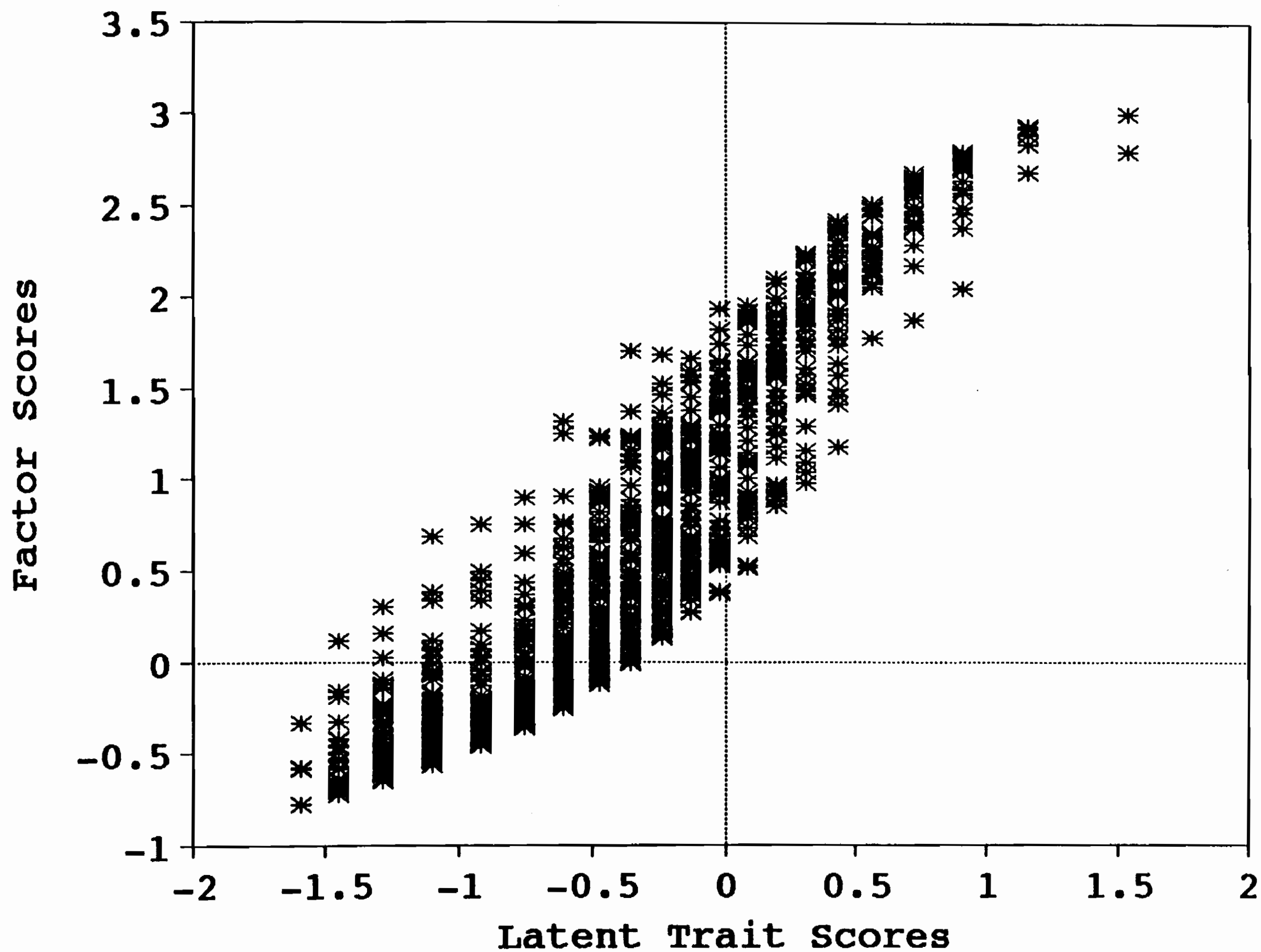


Figure 41. Scale 2 (latent trait by factor scores, L-3, dichotomous data).

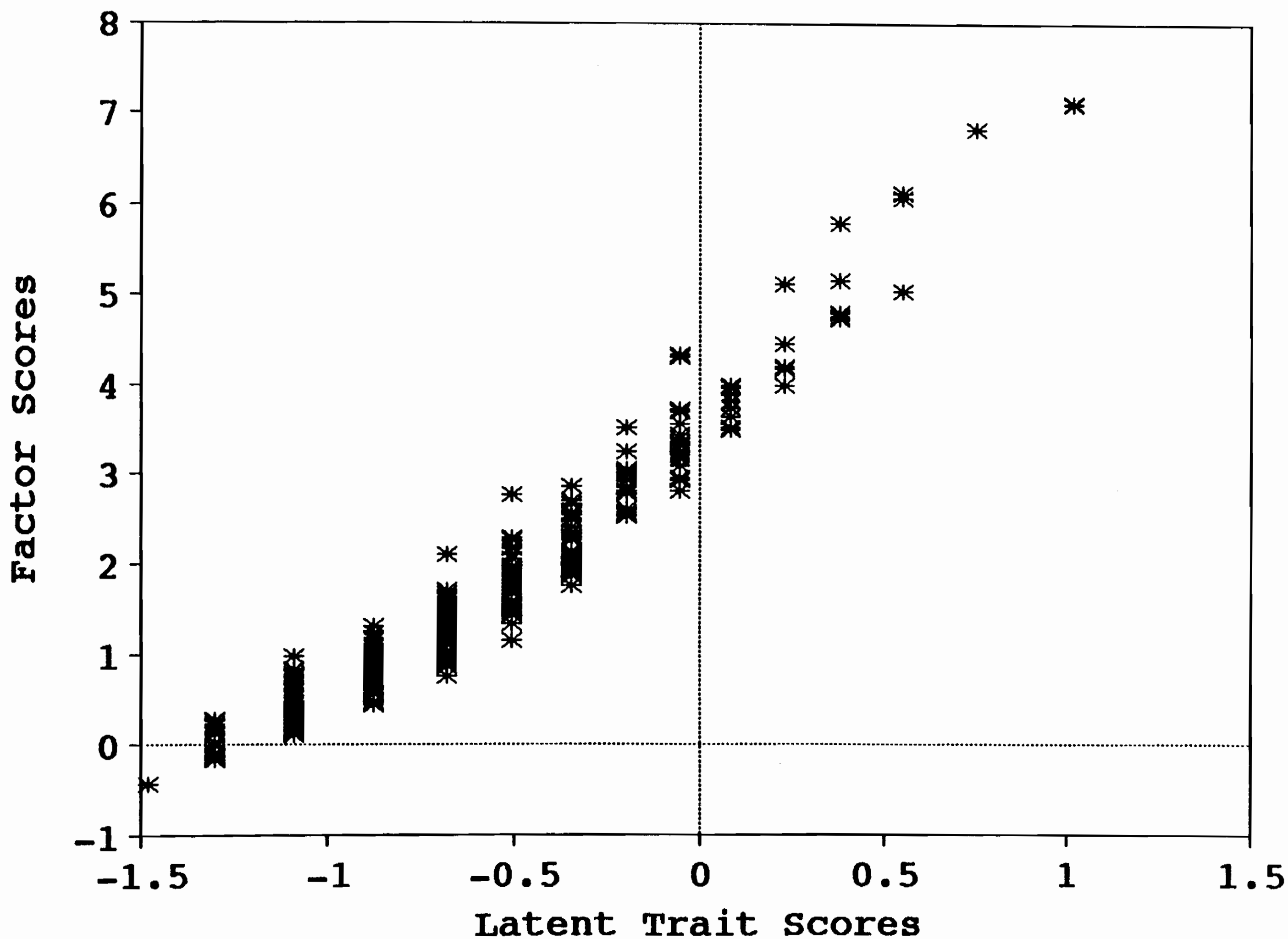


Figure 42. Scale 4 (latent trait by factor scores, L-3, dichotomous data).

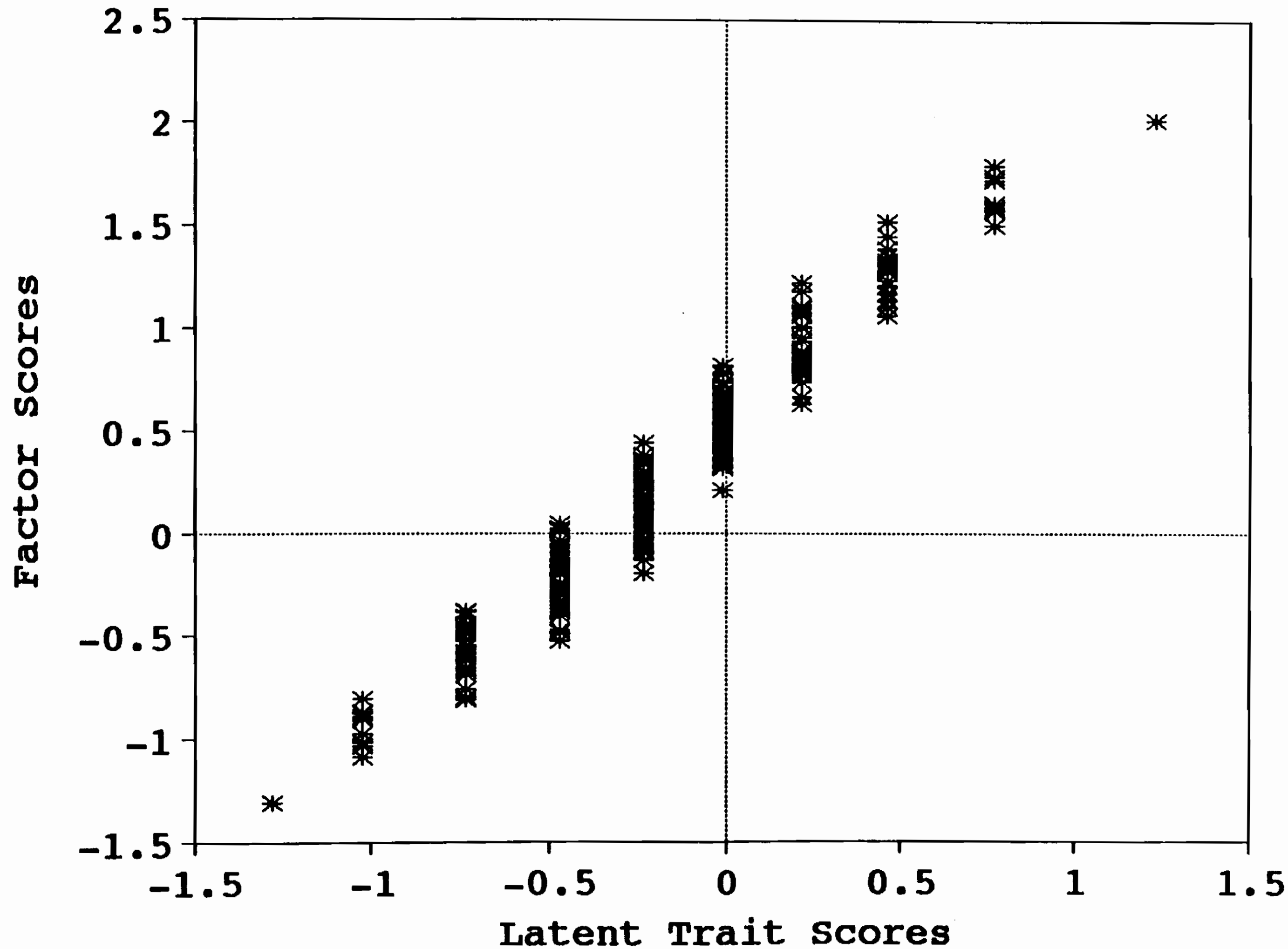


__Figure 43.__ Scale 5 (latent trait by factor scores, L-3, dichotomous data).

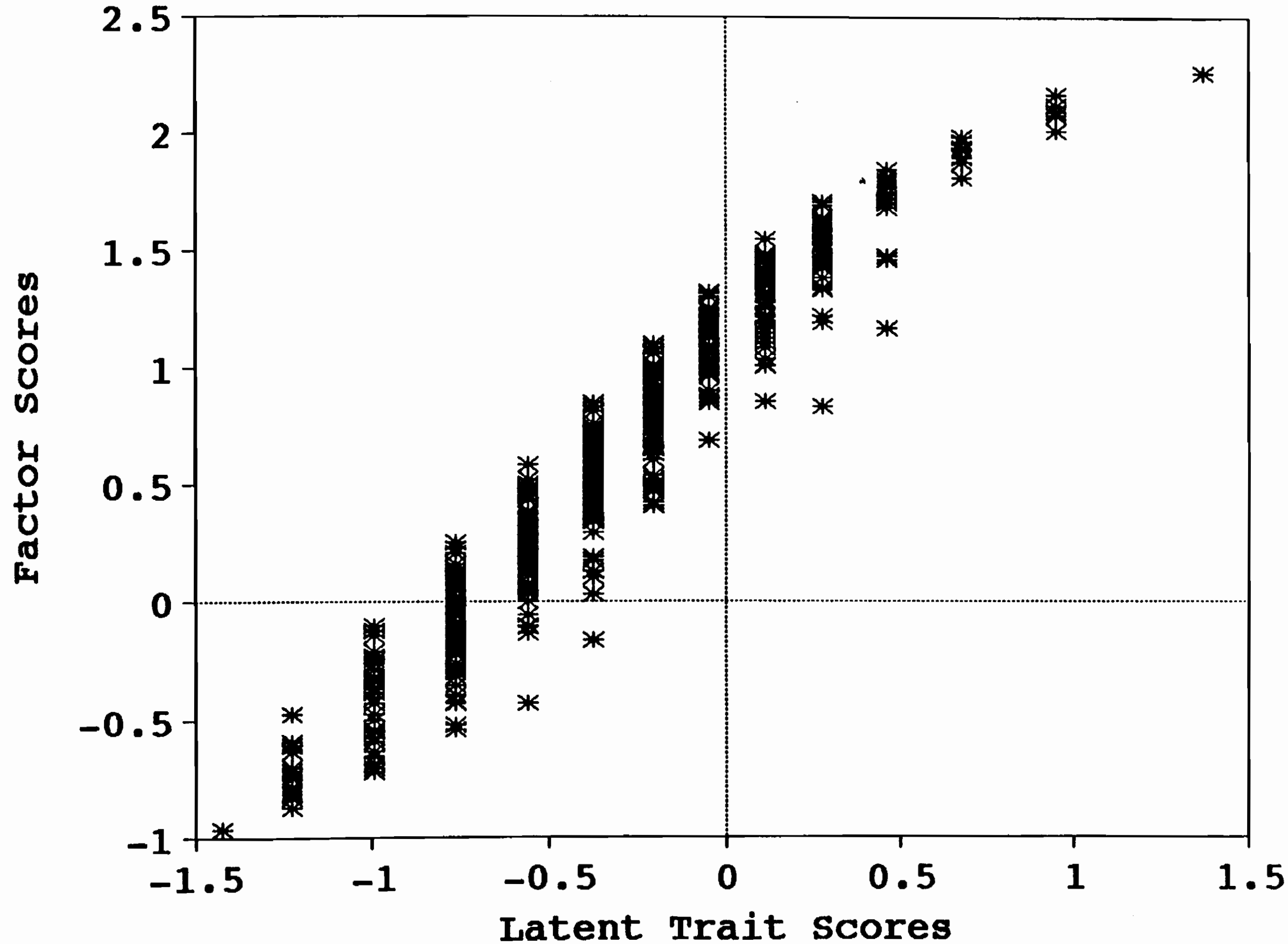


Figure 44. Scale 6 (latent trait by factor scores, L-3, dichotomous data).

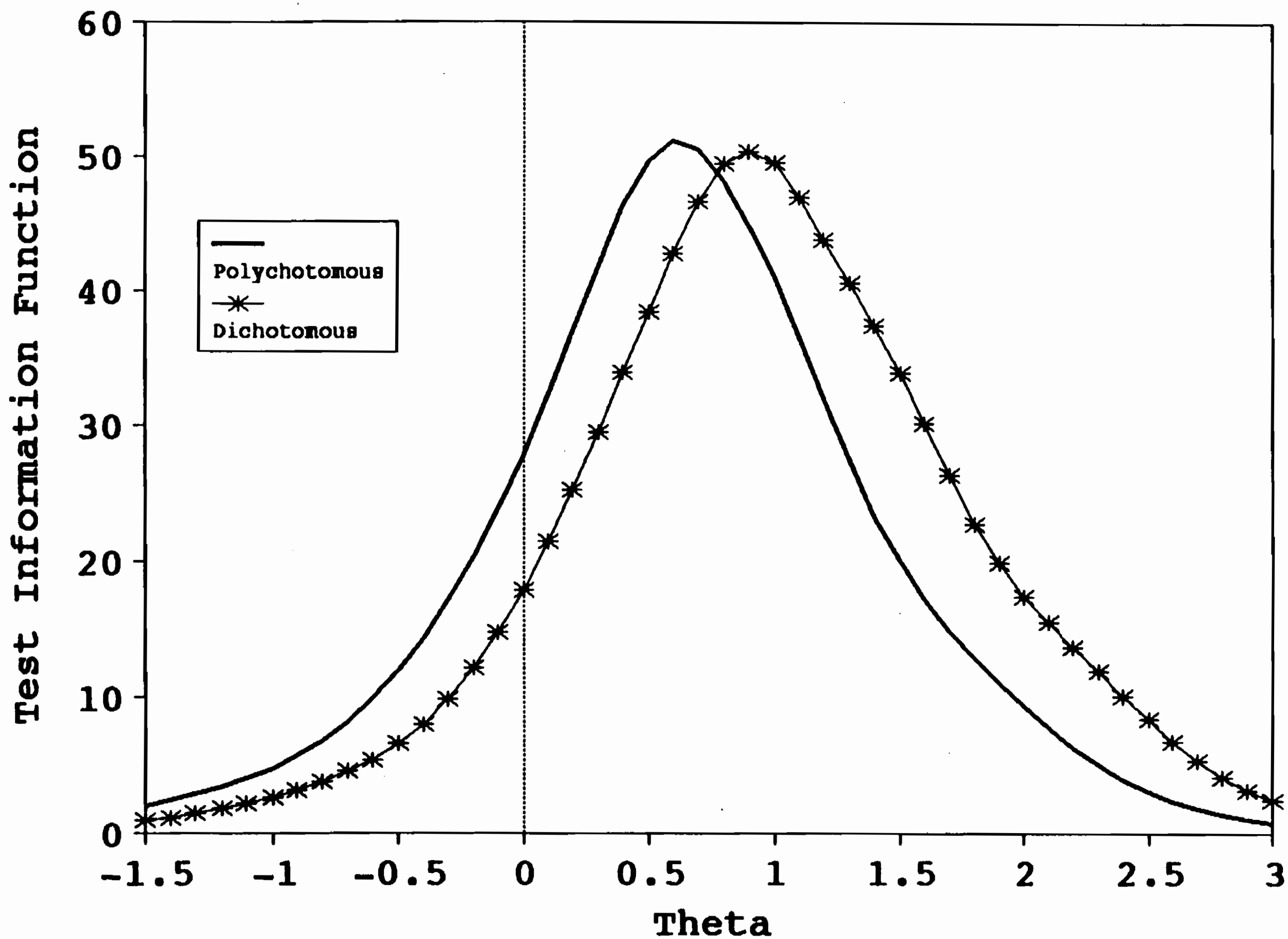


Figure 45. Test information function (scale two).

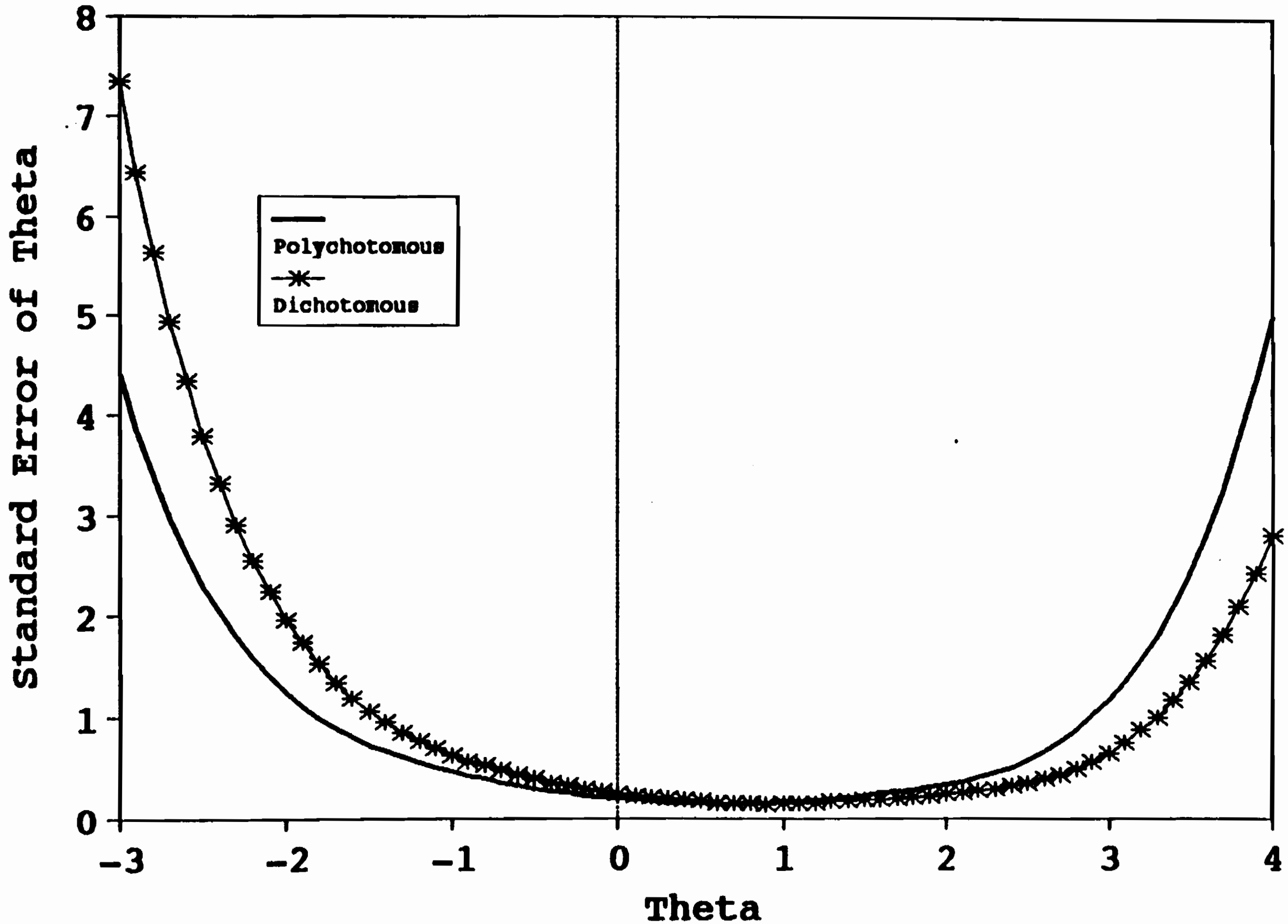
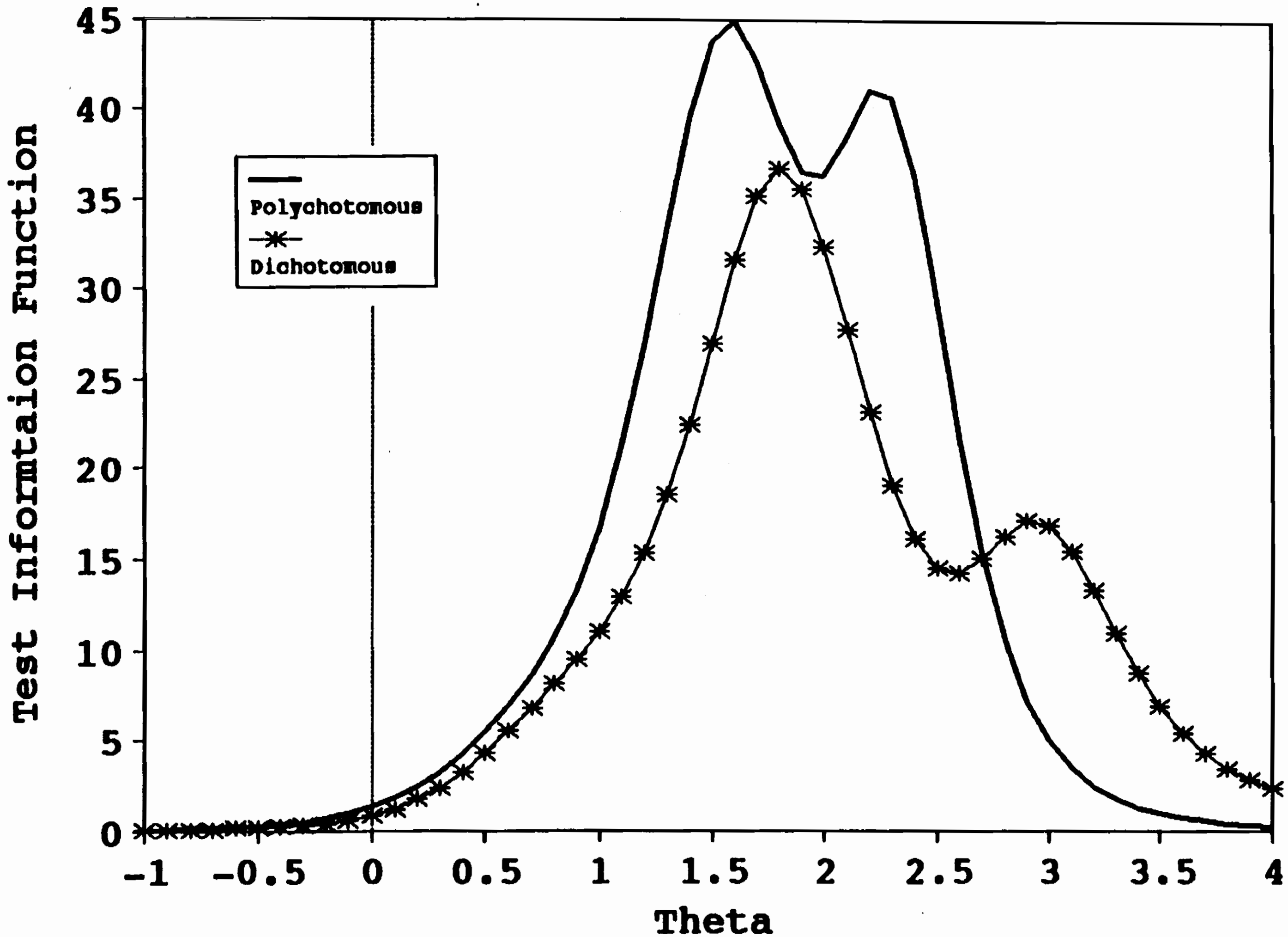


Figure 46. Standard error of theta (scale two).



<u>Figure 47</u>. **Test information function (scale four).**

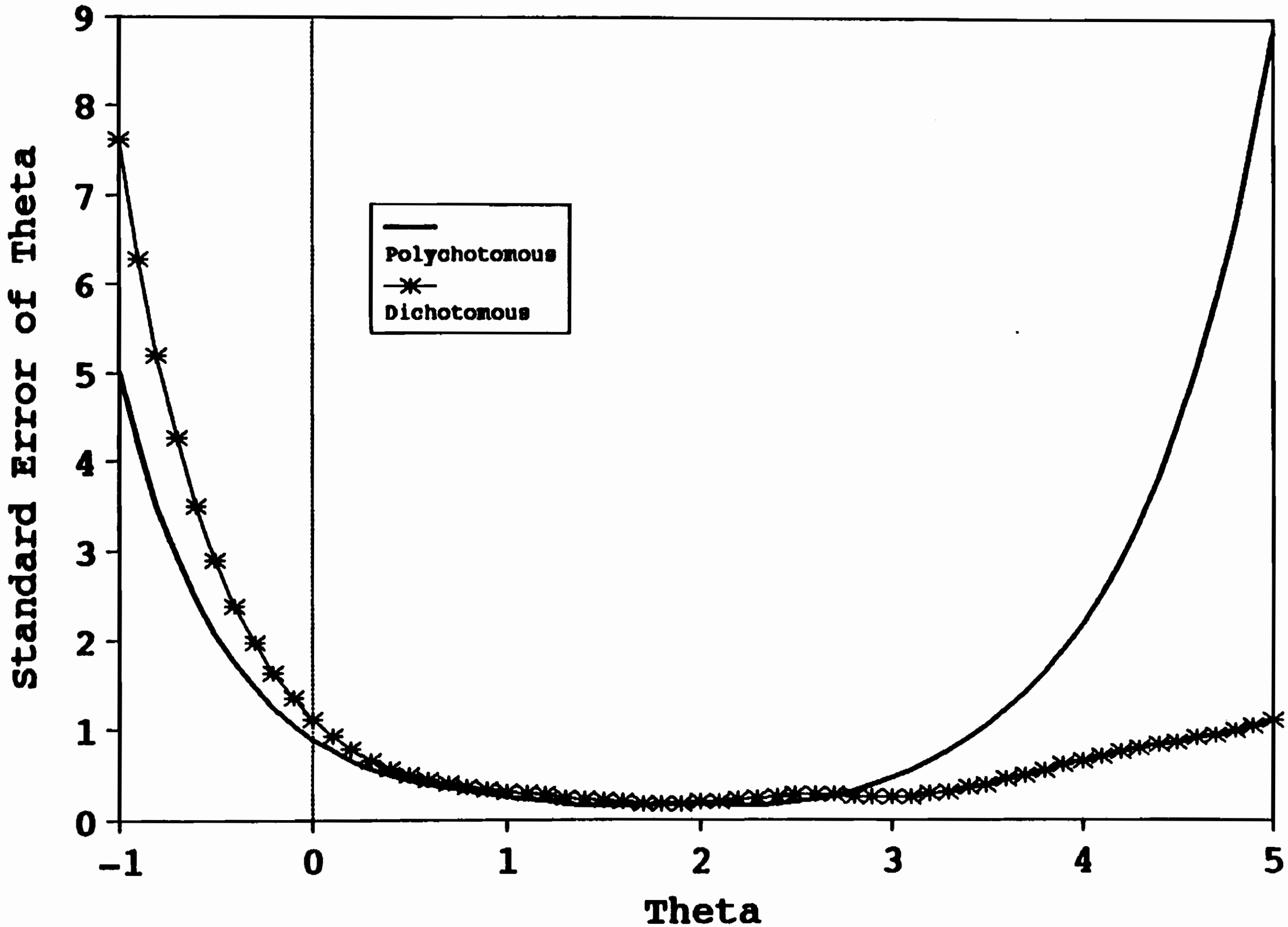
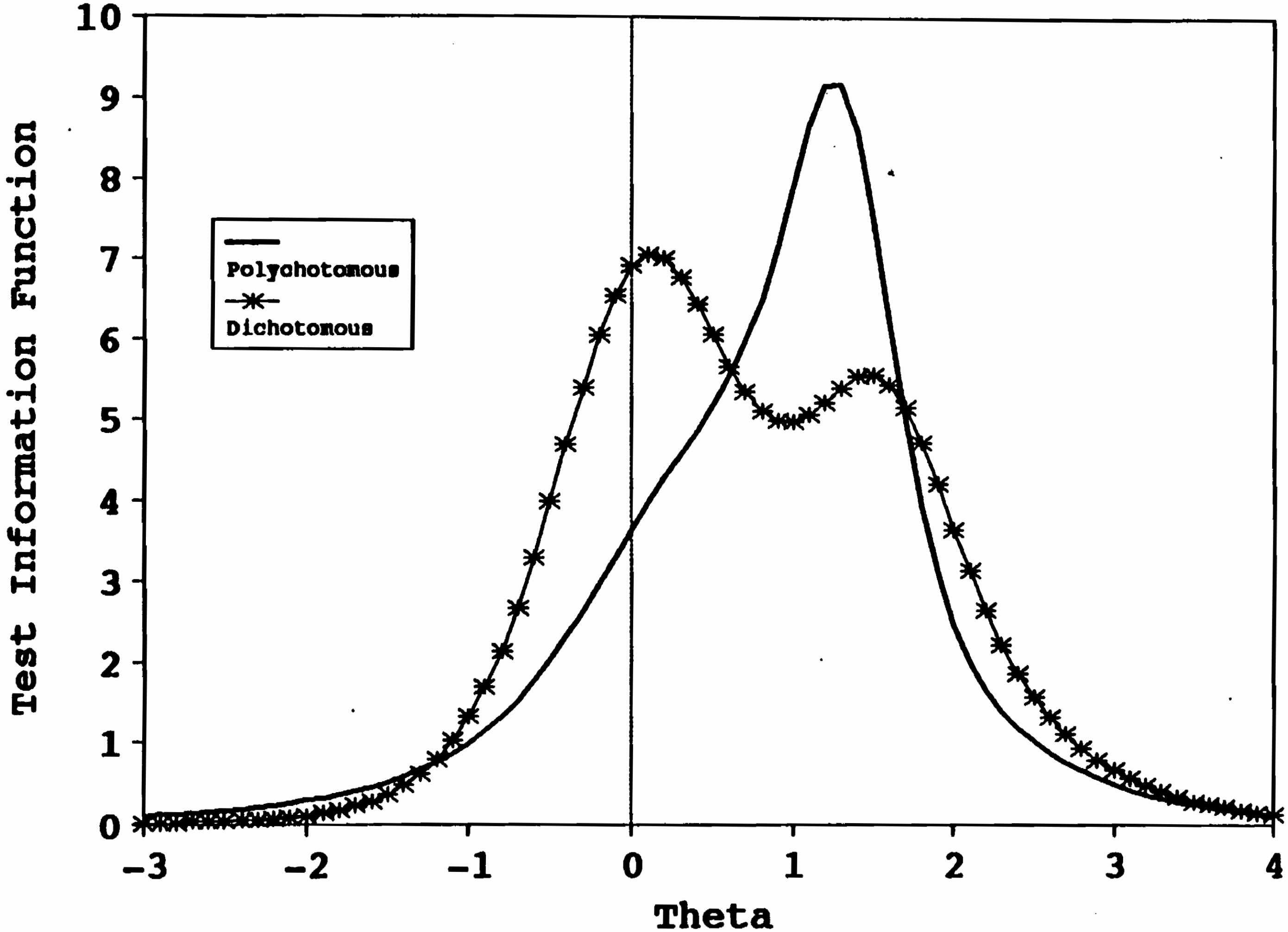
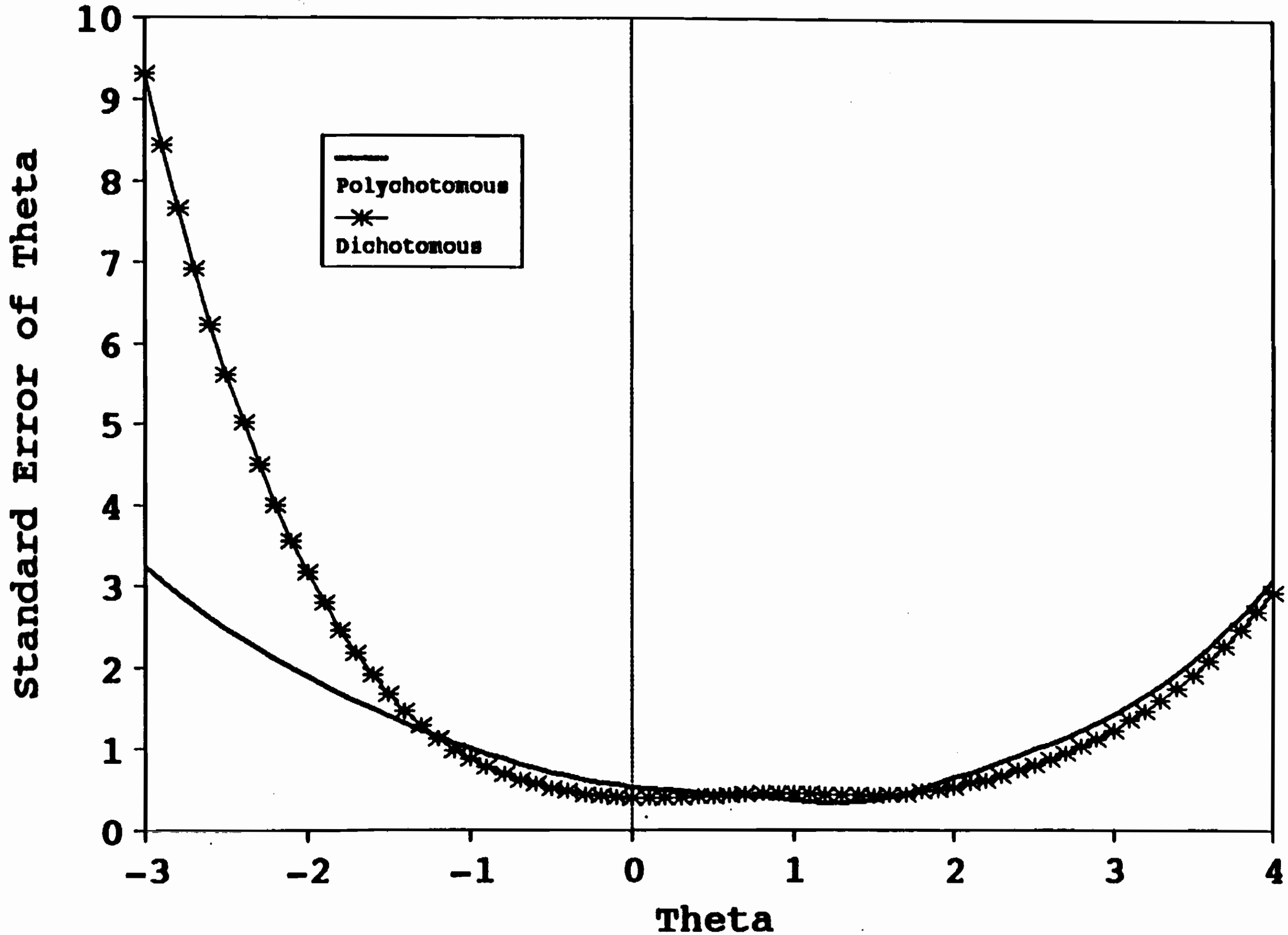


Figure 48. Standard error of theta (scale four).



<u>Figure 49</u>. Test information function (scale five).



<u>Figure 50</u>. **Standard error of theta (scale five).**

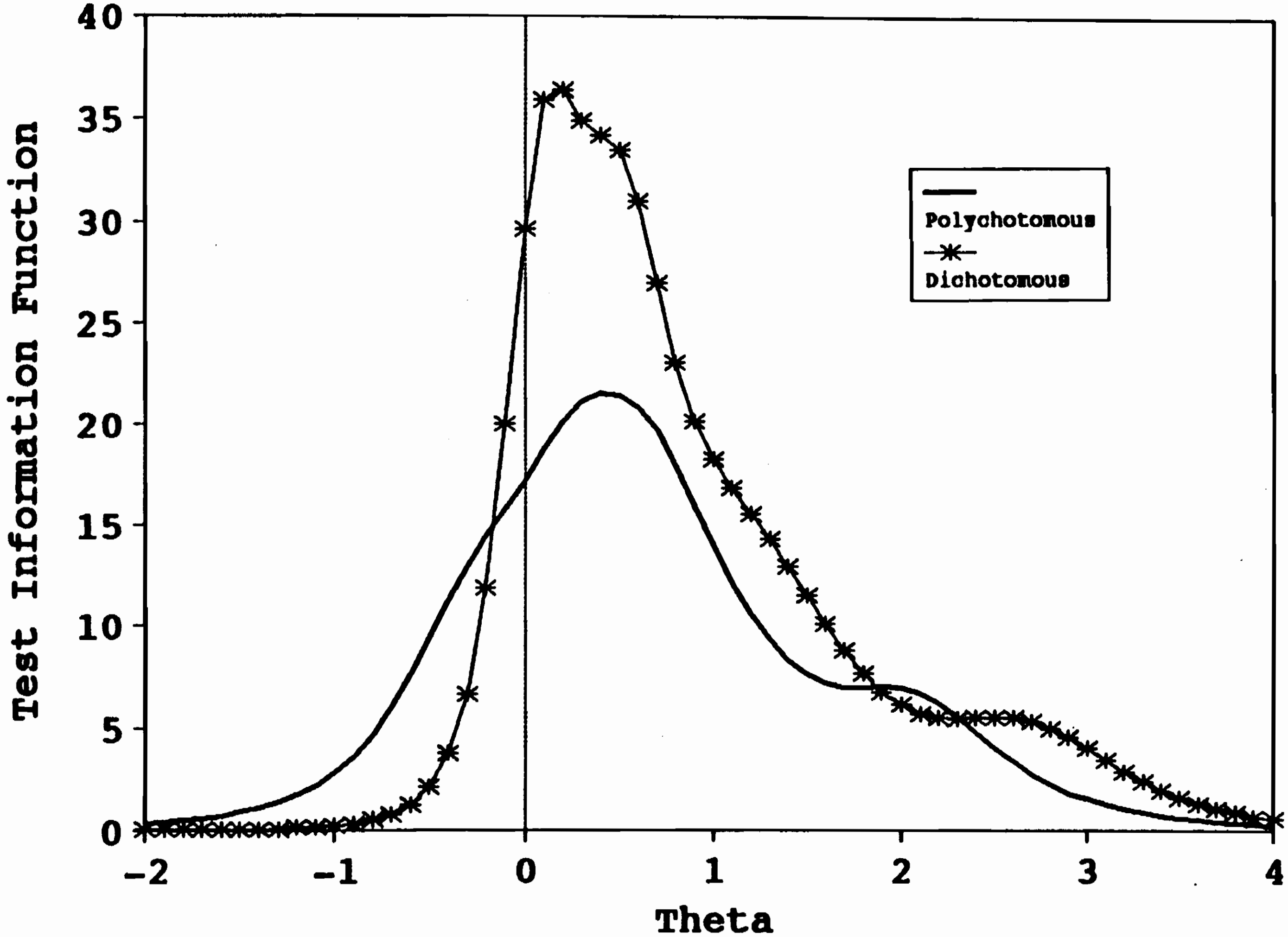
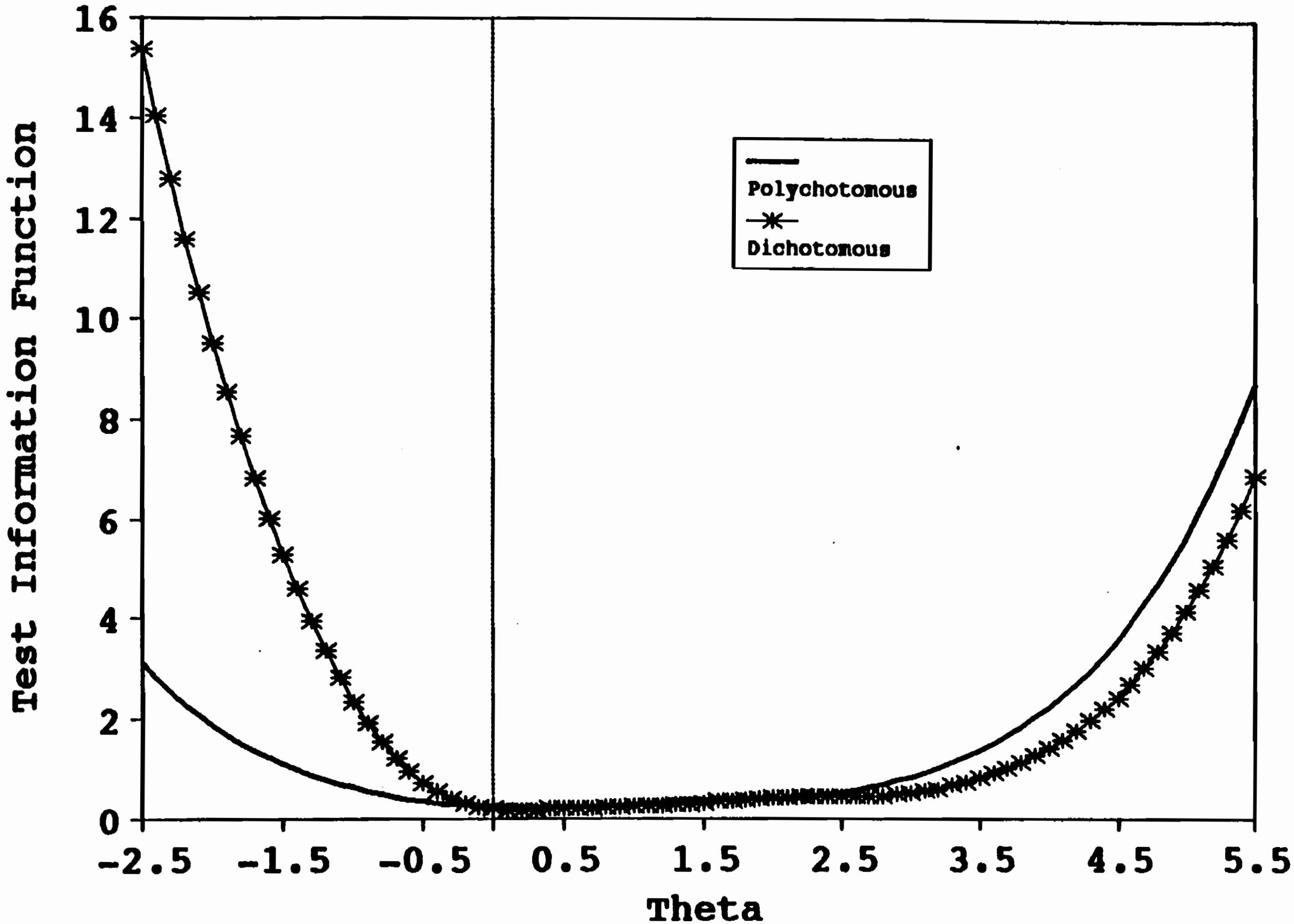


Figure 51. Test Information Function (scale six).



<u>Figure 52</u>. Standard Error of Theta (scale six).

VITA

R. Lance Becker

19919 Encino Briar
San Antonio, TX 78259
Telephone: 512-497-7643

Education

1973-1976	B.A., Psychology State University of New York at Stony Brook
1976-1978	M.P.S., Applied Clinical Psychology Long Island University at C.W. Post Center
1977-1978	Clinical Internship Sagamore Psychiatric Hospital
1987-1989	M.S., Industrial/Organizational Psychology Virginia Polytechnic Institute and State University. Thesis: Techniques for Identifying Dissimulation of the Managerial Potential Scale of the California Psychological Inventory
1989-1991	Ph.D. Industrial/Organizational Psychology Virginia Polytechnic Institute and State University (expected graduation - December, 1991). Dissertation: Latent Trait, Factor, and Number Endorsed Scoring of Dichotomous and Polychotomous Responses to the Common Metric Questionnaire

Employment

February, 1991- August, 1991	Senior Program Director The Psychological Corporation San Antonio, TX

Program manager for the division of professional assessment. Directed the development of test of assessment instruments for human resource functions. Worked with marketing, sales, production, and research and development departments to position products for use in employment situations. Developed and maintained productive relationships with test authors.

1989-1991 Industrial/Organizational Consultant (private practice)

Provided consultative services to organizations regarding a variety of human resource functions. Developed personnel selection, job evaluation, job specification, performance appraisal, program evaluation and validation strategies based on worker- and task-oriented job analyzes. Assisted human resource personnel in developing supervisor goal setting and communication training programs.

1989-1991 Instructor
Principles of Psychological Research

Fall, 1987 Graduate Teaching Assistant
Laboratory in Sensation and Perception

Spring, 1987 Graduate Teaching Assistant
Laboratory in Motivation and Perception

Fall, 1988 Assistant Coordinator, Introductory Psychology

Spring, 1988 Graduate Teaching Assistant
Laboratory in Advanced Social Psychology

1985-1987 Clinical Psychologist
Associated Mental Health Professionals
Staunton, Virginia

Provided individual, group, and family therapy to children and adults on an outpatient basis. Psychological, educational, and forensic assessments were performed. Consulted with community educators and school psychologists regarding special education and neuropsychological assessments. Biofeedback and relaxation training programs were conducted.

1984-1986 Program Director
Rockbridge Area Community Services Board
Lexington, Virginia

Developed a residential treatment program for children and adolescents. Supervised a staff of 15 professionals. Administrative responsibilities included personnel, budget, licensing, funding raising, and community relations. Developed and implemented a human resource information system for the Rockbridge Area Community Services Board.

1983-1986 Coordinator of Child and Adolescent Services
Rockbridge Area Community Services Board
Lexington, Virginia

Supervised a staff of 4 professionals in the treatment of children, adolescents, and families. A creative adolescent day treatment program was developed. Psychological and educational evaluations were provided. School based preventative mental health and drug abuse programs were developed and implemented. Assertiveness training, weight reduction, parent skill-building, and biofeedback and relaxation training programs were conducted.

1980-1983 Master's Level Clinical Psychologist
Rockbridge Area Community Services Board
Lexington, Virginia

Provided mental health services to children, adolescents, and adults. Psychological and educational evaluations were performed. Certified by the Virginia Supreme Court and the Department of Mental Health/Mental Retardation to conduct forensic evaluations. Crisis intervention services were performed for private and state hospital admissions.

1978-1980 Staff Psychologist
Conemaugh Valley Memorial Hospital
Community Mental Health Center
Johnstown, Pennsylvania

Responsibilities included psychological evaluations, behavioral assessments, and therapy with children and adults. Therapeutic treatment programs were developed in outpatient and residential settings. Therapeutic approaches included applied behavior analysis, strategic therapies, cognitive behavioral and supportive therapies. Two outreach centers were established. Liaison for the mental health center.

References

References furnished upon request.