CREATING A PARALLEL TEST FOR THE MYERS-BRIGGS TYPE INDICATOR USING ITEM RESPONSE THEORY

by
Leslie A. Thomas

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for

MASTER OF SCIENCE

in

Psychology

APPROVED:

Robert J. Harvey, Chairman

Neil M. Hauenstein

Sigrid B. Gustafson

October, 1994

Blacksburg, Virginia

CREATING A PARALLEL MYERS-BRIGGS TYPE INDICATOR USING ITEM RESPONSE THEORY

by
Leslie A. Thomas

Committee Chairperson: Robert J. Harvey
Psychology

(ABSTRACT)

A number of studies have concluded that the Myers-Briggs Type Indicator (MBTI) measures four general dimensions of personality; however, their findings also suggest that the MBTI's scales could benefit from improvements in their measurement precision. The degree to which the addition of newly constructed items to the Form F item pool would improve the measurement precision of the four primary MBTI scales was assessed. Using item response theory (IRT) to quantify each scale's performance, findings indicate that the new items substantially increased the test information functions (TIFs) and decreased the standard errors of measurement (SEM), especially in the critical area around the type cutoff scores (e.g., SEMs for scales containing the original plus new items were approximately half the size produced by the original MBTI items). The potential benefits of this increased measurement precision were discussed with respect to a number of applied testing issues.

Acknowledgments

Most of all, I would like to thank my committee chair, Robert J. Harvey. Without his ideas, expertise, and guidance (and unending patience), this research would not have been possible; I owe much of my graduate training to him. I would also like to thank my committee members, Sigrid B. Gustafson and Neil M. Hauenstein, for their assistance with this document. I thank Sigrid Gustafson for her many ideas, writing expertise, and contagious laughter (which is always a welcomed addition to a thesis defense). I would also like to thank Neil Hauenstein not only for being on my thesis committee, but also for assisting me with other research projects. He has often gone out of his way to support these projects and I will always greatly appreciate his efforts.

In addition, I would like to thank Gayle Kennedy, our secretary and much, much more. I could not have made it through many "last-minute" crises without her assistance. Finally, I would like to thank my family, in general, for their support and, in particular, my mother, Karen Thomas -- if it was not for her belief in me, I would not have made it this far: thank you.

TABLE OF CONTENTS

Overview of the Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI; Briggs & Myers, 1976) was designed for non-psychiatric populations as a means to operationalize and measure Carl Jung's theory of psychological types (1923; 1971). The MBTI is a paper-and-pencil personality inventory that utilizes a qualitative, forced-choice format in which examinees choose between two competing alternatives (e.g., word pairs, sentence completion format) to assess their standing on four bipolar dimensions of personal preference: introversion (I) versus extroversion (E); intuition (N) versus sensing (S); feeling (F) versus thinking (T); perceiving (P) versus judging (J). A numerical preference score is then computed for each of the four preference dimensions by summing the number of keyed item responses; these continuous preference scores are typically dichotomized to derive a categorical type assignment (e.g., E vs. I, N vs. S, T vs. F, and P vs. J) for each MBTI dimension.

Based upon their combination of four dichotomized scale scores, respondents may be assigned to one of sixteen possible personality types (Myers, 1962). The MBTI's authors developed narrative descriptions, or profiles, for each of the sixteen types that explain how each preference in a particular personality type will interact with the other preferences to influence an individual's behavior (McCaulley, 1990). The test's authors proposed that these profiles can be used to understand and predict behavior in personal and professional settings based on an individual's results on the MBTI (Myers & McCaulley, 1985).

Development of the MBTI

Kathleen Briggs and Isabel Briggs Myers developed the MBTI over a period of several decades based on the assumptions that preferences for Jungian types actually exist, these preferences can be measured via self report, and these preferences exist as dichotomies (i.e., have a bimodal distribution; Myers & McCaulley, 1985). The initial pool of questions was tested in a criterion group (i.e., individuals whose type was "known" via longitudinal observation by the test's authors) between 1942 to 1944, producing Forms A and B of the MBTI. These forms consisted of the same items ordered differently. Items that did not receive a response of 60% or more by examinees of the appropriate type were eliminated (Myers & McCaulley, 1985). A differential weighting scheme was also devised to account for the social desirableness of a particular response, with socially desirable items being given less weight than less popular items.

The next version, Form C, excluded any items that had a high validity for more than one scale of the MBTI. Form D was developed between 1956 and 1958, starting with a pool of 200 new items -- including the addition of word pairs -- in preparation for a new edition of the manual under the guidance of Educational Testing Services (Stricker & Ross, 1964). Old and new items were put into Forms E and F. Form F was published in 1962 by the Educational Testing Service and is used in large scale data collections; it is also the form recommended for research (Myers & McCaulley, 1985). Form G, published in the late 1970s, is a shorter version of Form F (126 vs. 166 items) with many of the experimental and ambiguously worded items removed. Form AV (Abbreviated Version)

or H, is an abbreviated version of Form G with only 50 items, is self-scoring, and is designed for group situations (Murray, 1990).

The MBTI as a Test of Jung's Psychological Types

As a measure of Jungian theory (1923; 1971), the goal of MBTI is to categorize people according to their psychological type. A type is a "category into which people with similar but not necessarily identical characteristics are placed" (Hall & Nordby, 1973, p.96). Categorizing people based on their psychological type is based upon the Jungian assumption that much of seemingly random human behavior is actually quite orderly and consistent if viewed in terms of what Jung termed "preferences".

A preference is "one of four basic dichotomies that in type theory structure an individual's personality" (e.g., E-I, S-N, T-F, and J-P dichotomies; Myers & McCaulley, 1985, p 225). Preferences are thought to be primarily developmental in nature; that is, over time, interactions between individuals and their environment establish the relative predominance of one preference over the other, although the capacity to use either preference always exits.

These preferences are thought to reflect different aspects of individual differences in two major mental processes: perception and judgment -- variations in the styles by which individuals gather information and then reach decisions based upon that information. These two processes are thought to be important because in every situation; an individual must first perceive and then make a judgment based upon those perceptions. (Myers & McCaulley, 1985).

To describe variations in the perceptual process, Jung used the sensation and intuition (S-N) bipolar dimension to reflect two different modes of perceiving. The sensing type (S) prefers the perception of what is directly observable through the five senses. They are often described as realistic, practical, conservative, and experiential. The intuitive type (N), on the other hand, prefers to extrapolate from the information given by the senses and look for meanings, possibilities, and relationships on a more abstract or conceptual level. Intuitive types are often described as artistic, analytical, creative, liberal, and theoretical.

Once data has been collected through the perceptual process, a conclusion is drawn based on that data (Myers & Myers, 1980). The thinking and feeling (T-F) bipolar dimension reflects two different ways this judgment can be made. The thinking type (T) prefers to arrive at a judgment by logical and impersonal methods in contrast to the feeling type (F) who tends to be more subjective and bases decisions primarily on personal values. Thinking types are generally described as dominant, assertive, achievement-oriented, and autonomous while feeling types purportedly have high needs for affiliation, warmth, and harmony (Thompson & Borrello, 1986).

The last bipolar dimension, the judging and perceiving (J-P) dimension, represents an <u>overall</u> preference for using either the <u>perceiving</u> function (S-N) or <u>judging</u> function (T-F) in everyday interactions with the environment (DeVito, 1985). Unlike the other dimensions measured by the MBTI, the J-P dimension was implied but never explicitly defined by Jung as an independent personality dimension (Carlyn, 1977; Myers &

McCaulley, 1985). Judging types prefer things to be planned, orderly, completed, and issues to be resolved. Perceiving types, on the other hand, prefer to live in a more flexible, spontaneous, and adaptive manner, generally prefer to keep as many options open to them as possible and, as a result, do not like to make decisions because they view it as limiting their options (e.g., such a setting an appointment to meet someone at a certain time or working according to a predetermined schedule; Myers & McCaulley, 1985).

Although though the test was developed based on Jungian type theory, most researchers agree that the MBTI measures only restricted aspects of the Jungian dimensions it was intended to reflect (Coan, 1978; Comrey, 1883; McCrae & Costa, 1989; Murray, 1990; Sipps & DiCaudo, 1980; Stricker & Ross, 1964). For instance, its measures of the E-I dimension often focus on only observable behavior such as social extraversion-introversion, a dimension of gregariousness in contrast with shyness and withdrawal (Palmiere, 1972). Critics argue that assessments of only external behavior probably cannot be sensitive enough to measure the complicated Jungian personality schema and, consequently, will sacrifice some of the theoretical richness of Jungian insights (Carlson, 1980; Coan, 1978; Cowan, 1989; McCrae & Costa, 1989; Steele & Kelly, 1976).

Moreover, many argue that since preferences are assumed to interact in complex nonlinear ways to produce psychological types (Murray, 1990), the ideal assessment procedure for Jung's typology would take the pattern of each component's interaction into account rather than simply assessing which component is preferred in various situations as

does the MBTI (Coan, 1978). As one critic derisively put it the "Myers-Briggs stereotypes people, it is a *static, undynamic* [italics added] theory that traffics labels much like astrology" (Moore, 1987, p. 82).

But according to Murray (1990) and other researchers, Jungian type theory is probably "too complicated to be captured on an objective personality test" (p. 1195). For instance, much of the theory is vague and ill-defined, often dealing with archetypes and images not easily communicated -- much less assessed -- via an objective test (Carlson, 1985). In addition, much of Jung's theory deals with the unconscious which, by definition, can not be assessed via direct self-report. Furthermore, some of Jung's trait-like descriptions of the four dimensions (e.g., his description of extraverts as being open, jovial, social, and approachable, as well as morally conventional and tough-minded) define observable characteristics that do not empirically covary; that is, they do not define the <u>same</u> underlying construct/dimension (Guilford, 1977).

In fact, some of the descriptions for various types contained in the MBTI directly <u>contradict</u> Jung's writings on psychological type theory (McCrae & Costa, 1989). For example, according to Jung's theory, a feeling person is one who assigns a value (either positive or negative) to an experience or object without a logical basis. However, an individual who assigned a negative value to other people such as hatred or mistrust, without a rational reason, would tend to score very low on the feeling dimension of the MBTI which is defined by illustrative adjectives such as compassionate, warm, and trusting (Coan, 1974).

Although the test is not an isomorphic translation of Jungian type theory, many researchers, while keeping that caveat in mind, still contend that the instrument has value as "a practical assessment instrument whose constructs have been clarified by extensive research" (Murray, 1990, p. 1195). McCrae and Costa (1989) take this argument a step further by arguing that "the MBTI does not seem to be promising instrument for measuring Jung's types; those who embrace Jung's theory should probably avoid the MBTI" (p. 32). Instead, they propose that it might be better to abandon the Jungian theoretical framework altogether and adopt the more comprehensive and accepted personality taxonomy of the five-factor model.

The MBTI and the "Big Five"

McCrae and Costa's (1989) contention that users of the MBTI would benefit from a "radical reinterpretation" of the MBTI from a five-factor perspective is based on their findings that the MBTI dimensions correlate significantly with four of the five Big Five dimensions - Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Using the NEO-PI (Costa & McCrae, 1985) as their Big Five measure, they found that the MBTI's extraversion pole of the E-I dimension was significantly related to the Big Five's Extraversion dimension, the intuition pole of the S-N dimension was significantly related to the Big Five's Openness to Experience dimension, the feeling pole of the T-F dimension was significantly related to the Big Five's Agreeableness dimension, and the judging pole of the J-P dimension was significantly related to the Big Five's Conscientiousness dimension.

Not surprisingly, given the positive and affirming nature of MBTI (Saunders, 1986), McCrae and Costa (1989) did not find a relationship between any of the four MBTI dimensions and the Neuroticism/Emotional Stability/Adjustment dimension of the five-factor model. They concluded that although these findings are in keeping with the MBTI author's intentions, the absence of such a dimension "omits information that may be crucial to employers, co-workers, counselors, and the individuals themselves. For many, if not most, applications, some measure of Neuroticism would be useful" (p. 36).

However, Harvey, Murry, and Markham (1994) demonstrated that the 71 "unscored" experimental items contained in Form F of the MBTI (only 95 of the 166 items are scored to produce the four bipolar dimensions; Sipps, Alexander, & Friedt, 1985) could be scored to produce a fifth dimension that corresponds to the Big Five Neuroticism dimension. The correspondence of the item content between the five MBTI dimensions (including the "new" Neuroticism dimension) and the Big Five dimensions "indicates a solid match between the 5-factor MBTI solution and the conceptual definitions of the Big Five's constructs" (Harvey et al., 1994; p. 9). These findings are particularly important because they give practitioners the ability to interpret MBTI scores based on either traditional interpretations (i.e., Jungian type theory) or within a Big Five framework.

The Growing Use of Personality Tests in Industry

Regardless of whether practitioners decide to adopt a big-five reinterpretation of the MBTI or stay with its original interpretation in the future, at present, the MBTI is the most extensively used personality inventory in American industry (Zemke, 1992). In

addition, its recent use in industry has experienced the fastest growth rate of any personality inventory; in 1986, US. businesses accounted for 40% of test sales -- twice that of their share of just three years previous to that time (Moore, 1987). In 1991 alone, it was estimated that over 2 million people took the MBTI (Moore, 1987; Suplee, 1991; Zemke, 1992); a substantial increase from the approximately 1.5 million that completed the test in 1986.

The MBTI's extensive use in industry is indicative of the recent resurgence in popularity experienced by personality tests in general. The prevailing view from the 1960's until recently was that personality tests were poor predictors of job performance (e.g., Guion & Gottier, 1965; Hogan, 1991). However, recent developments have given way to a more optimistic view of their use in industry. First, evidence via recent meta-analytic and empirical studies have supported the utility of personality measures as predictors of diverse performance-related criteria (Barrick & Mount, 1991; Cortina et al, 1992; Day & Silverman, 1989; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Matthews et al., 1992; Mount et al., 1994; Schmitt & Ryan, 1993; Tett et al., 1991).

Second, it has also been demonstrated that some personality measures can provide substantial incremental validities over cognitive measures for the prediction of a variety of job-related criteria (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). For instance, personality inventories can complement the use of cognitive tests by tapping such dimensions as an individual's willingness to learn (e.g., Openness to Experience) in addition to the domains usually measured by cognitive tests such as the ability to learn.

The measuring of noncognitive factors, such as an individual's motivation to learn, is important because it is generally agreed that noncognitive factors are heavily implicated in many, if not most, aspects of job performance (Barrick & Mount, 1991; Hogan, 1991).

Third, recent developments in the legal environment with respect to selection and testing have made personality testing a viable alternative to the more traditional types of testing. With the passing of the 1991 Civil Rights Act, employers can no longer adjust test scores or employ different standards for testing on the basis of minority status in order to avoid adverse impact (e.g., race norming). In addition, the shifting burden of proof model provision should make it easier for plaintiffs to win disparate impact cases; therefore it is likely that the number of discrimination cases will increase (Pattison & Varca, 1991). As a result, employers may utilize more personality-based selection measures because, unlike most cognitive measures, they tend to have little, if any, differential impact on protected groups, and are therefore less prone to raise discriminatory concerns (Hogan, 1991). Finally, with the passing of the Employee Polygraph Protection Act of 1988, paper-and-pencil personality tests -- both clear-purpose (i.e., directly asks information about attitudes toward theft) and general-purpose type tests (e.g., an employee reliability index derived from the Hogan Personality Inventory (HPI); Hogan & Hogan, 1988) -- have replaced the traditional "lie detector" as the new means of assessing honesty in the workplace (Gatewood & Field, 1990).

Industrial Applications of the MBTI: Understanding vs. Predicting Behavior

Although much of the MBTI's current attention may be viewed as a mere by-product of the recent growth in popularity of personality tests in industry, in general (Moore, 1987), a review of the literature citing the various ways in which it has been utilized (e.g., selection, job placement, and performance appraisals) suggests that corporations view an employee's MBTI profile as anything but a passing fad. The MBTI has become an integral part of a substantial number of management development programs throughout corporate America (Moore, 1987). Corporations such as American Telephone and Telegraph, General Electric, Exxon and Transamerica use the MBTI profiles of employees to improve decision-making processes and enhance team-building efforts (Coe, 1992; McCrae & Costa, 1989). Some of the more frequent applications include resolving employee disputes, enhancing interpersonal and inter group relations (Coe, 1992; Gauld & Sink, 1985; Moore, 1987), performance appraisals (Agor, 1988), strengthening customer relations (Moore, 1987), training (Coe, 1992), and vocational counseling (Poilitt, 1982).

Such applications within businesses rest upon the assumption that knowing each other's MBTI profile facilitates communication between members and helps them understand themselves better in relation to their work and the other members within the organization (Cowan, 1989). Organizations use the MBTI because employers believe that this increase in understanding among members will lead to increased motivation,

problem-solving, innovation and overall productivity within the organization (Hoy & Hellriegel, 1982).

Research indicates that the MBTI has also been used primarily as a predictive instrument. For instance, MBTI profiles have been used to predict employee turnover (Garden, 1989), various aspects of job performance (Kirton, 1976; Gough, 1976), organizational consulting competence (Bushe & Gibbs, 1990), as well as the cost efficiency of an employee's decision-making style as assessed by the MBTI (Davis, Grove, Knowles, 1990). The MBTI has also been used for selection and promotion purposes with corporations trying to match the characteristics of a candidate's MBTI profile to the characteristics of the job or of successful job incumbents (Coe, 1992; Zemke, 1992). Job analysis has been used in conjunction with this latter application to obtain estimates of the MBTI profiles most likely to be found among job incumbents based upon the scoring of the Position Analysis Questionnaire (PAQ; McCormick, Jeanneret, & Meecham, 1972) for a particular job.

Psychometric Issues

Surprisingly, little has been done with the MBTI in terms of rigorous psychometric research (Cowan, 1989). The practical implications of some of the applications of the MBTI -- especially selection, placement, and promotion -- underscore the importance of achieving adequate psychometric properties (e.g., reliability and validity), as well as demonstrating its practical utility in the situations for which it is being recommended and/or used (Zemke, 1992). The most basic of these issues concerns the MBTI's

reliability, given the fact that reliability is a prerequisite for both validity and utility (Crocker & Algina, 1986); that is, a test that does not consistently (reliably) measure a phenomenon cannot, therefore, effectively predict the phenomenon and, without prediction, a test cannot be useful.

Reliability

Although a number of studies have concluded that the MBTI demonstrates satisfactory test-retest, split-half, and internal consistency reliabilities, with estimates ranging from the low .70's to low .90's (Carlson, 1985; Carlyn, 1977; Carskadon, 1982; DeVito, 1985; McCarley & Carskadon, 1983; Myers & McCaulley, 1985; Tzeng, Outcalt, Boyer, & Ware, 1984), the question of what constitutes a satisfactory reliability level ultimately depends upon one's purpose for using the test, and the way in which the test is scored.

Specifically, if the purpose for using the test is to predict future behavior in a situation in which erroneous decisions could be detrimental to both individuals and the organization (e.g., a selection or promotion decision), then a high degree of reliability is desirable. The need for a higher degree of reliability is compounded by the fact that the scoring system of the MBTI utilizes a cutoff score to dichotomize the continuous preference score when assigning categorical type labels. Clearly, the relative importance of measurement error increases dramatically in situations in which a cutting score is used to divide the distribution into discrete types because the effect of unreliability in the

vicinity of the cutoff score could cause an examinee to be assigned to the opposite preference category (e.g. Extravert instead of an Introvert).

To illustrate the need for increased measurement precision for the MBTI scales, Harvey and Murry (1994) -- used a pooled MBTI preference score reliability of .85 to set confidence intervals around the preference score cutoff points, the goal being to determine what percentage of examinees would fall within the measurement error band (i.e., plus or minus 1-2 standard errors of measurement) around each cutoff score. Their results indicated that approximately 22-55% of examinees fell within the area of measurement imprecision around the cutoff scores; for such individuals, even a slight amount of measurement error could cause them to be classified in the opposite type category. The consequences of being misclassified on just one dimension are exacerbated by the proposed interactive nature of the four dimensions (Myers & McCaulley, 1985), in that a change on only one dimension could drastically alter the overall interpretation of one's MBTI profile (and therefore any predictions based upon it). For example, an ENTP profile is interpretively quite different from an INTP profile, not just in terms of the E-I dimension but also how the other three dimensions interact with the E-I dimension.

A related matter concerns test-retest findings regarding type stability, or the probability of being assigned the same categorical personality type on repeated testings. If a test is adequately reliable, it should yield similar on repeated administrations; or, in the case of the MBTI, it should assign individuals to same categorical personality type on repeated administrations (barring, of course, the possibility of true type change during the

inter-test interval [ITI]). On a practical level, this aspect of the MBTI is most important if employee selection predictions are to be made based on the categorical type profiles (a common practice for the MBTI).

Although McCarley and Carskadon (1983) similarly concluded that the test-retest reliability coefficients of the MBTI subscales within a five week retest period were "high," ranging from .77 (T-F scale) to .89 (JP scale), the percentage of subjects who retained their dichotomous type preferences across all four scales was only 47% (Carlson, 1985). Thus, respondents had approximately a 1-in-2 chance of being assigned the opposite preference category on at least one of the MBTI personality dimensions upon the next test administration. Carskadon (1979) reported that test-retest reliabilities over five-week intervals ranged from .78 to .87; analysis of type stability revealed that 19% of the subjects having changed type on the E-I dimension, 11% on the S-N dimension, 17% on the T-F dimension, and 16% on the J-P dimension. The short test-retest ITI in both of these studies suggests that this type-shifting phenomenon is not developmental in nature; indeed, given the short ITI, it is likely that the majority of these apparent type changes were due simply to the action of measurement error in the vicinity of the type cutoff scores.

In 1991, the National Research Council (NRC), a subgroup of the National Academy of Sciences, examined the MBTI and five other instruments used extensively in training (Zemke, 1992). After reviewing 11 studies of MBTI test-retest outcomes, the NRC found that type stability ranged anywhere from 24 percent to 61 percent. In other words, out of 100 examinees, at least 24 and at most 61 of them would be assigned to the

same type category when retested over various retest intervals. The median for those changing on at least one of the four dimensions was 37%, with a low and high of 27 and 44 percent, respectively.

Even the type stability statistics for various ITIs documented in the MBTI manual (Myers & McCaulley, 1985) demonstrate clearly less than perfect test-retest agreement for type categories. For a 5-week ITI, the percentage of categories that remained unchanged on retest for all four dimensions (e.g., the chance of an ENTP staying categorized as an ENTP upon the second test administration) ranged from 43% to 56% across four independent samples. In one study using a 5-week ITI, the percentage of test-retest agreement in each category ranged from 54% to 100% for the E-I dimension, from 69% to 100% for the S-N dimension, from 67% to 100% for the T-F dimension, and from 58% to 100% on the J-P dimension. These statistics demonstrate that for each dimension some individuals remained the same type (100% agreement) while others had anywhere from a 46% chance (for the E-I dimension) to a 31% chance (for the S-N dimension) of changing on at least one of the dimensions. For longer ITIs, varying from 2 months to 2 years, the type stability statistics were even more dismal: only 31% to 53% of the subjects stayed the same four-letter type, with a median percentage of 39%, across eight different studies.

One possible reason for this discrepancy between apparently adequate test reliability statistics and the clearly less than desirable type stability statistics is that many of the test-retest and split-half reliabilities reported for the MBTI may be somewhat overestimated. That is, almost all of the studies that have assessed MBTI reliabilities used

the continuous preference scores to compute the reliabilities (Carlson, 1985; Murray, 1990). Using continuous scores rather than the dichotomous type scores that are actually used to define the categorical type profile might tend to inflate the reliabilities, because some degree of information loss would be expected as a result of the dichotomization process. The loss of information caused by dichotomizing the preference score may be substantial; Harvey and Murry (1994) estimated that from 26-32% of the preference score variance was lost due to dichotomization. As a result, reliabilities based on continuous scores may be somewhat, if not significantly, higher than those based on dichotomous preference scores. Since interpretations and predictions are ultimately made on the basis of dichotomous classifications (Myers & McCaulley, 1985), it could be argued that the reported reliabilities should be also be a reflection of these dichotomies.

Unreliability in an instrument, such as the MBTI, may be a function of poor item sampling; that is, if items that do not measure the desired construct well are present, this will introduce construct-irrelevant or error variance, thereby decreasing the precision -- and therefore the consistency -- of the instrument (Crocker & Algina, 1986). Numerous studies (Comrey, 1983; Johnson & Saunders, 1990; Sipps, Alexander, & Friedt, 1985; Thompson & Borrello, 1986, 1989) have demonstrated that MBTI items vary greatly in terms of the strength of their factor loadings, and that some items are much better measures of the underlying MBTI constructs than others (Comrey, 1983).

Classical Test Theory versus Item Response Theory

Most item analysis of the MBTI have been based on classical test theory (CTT) and, therefore, on the assumption of a linear relationship between a respondent's ratings of the MBTI items and the underlying personality construct (e.g., true score = observed score + error or the expected observed score). If this assumption of linearity is incorrect -- as may often be the case in applied measurement situations (Hambleton, Swaminathan, & Rogers, 1991; Hulin, Drasgow, & Parsons, 1983) -- then item statistics that are based on a nonlinear relationship between items and latent traits might more accurately reflect the item's functioning in relation to the underlying construct being measured. Unlike CTT, latent-trait or item response theory (IRT) models assume that the underlying trait being measured (termed theta) will demonstrate a nonlinear, S-shaped relation (term the item characteristic curve, or ICC) with the observed item response.

These ICCs underscore a fundamental difference between IRT and CTT: the way in which IRT links the observed item response to the underlying trait being measured. Specifically, it is assumed that the probability a correct response to an item is a function of both theta (the latent trait) and the characteristics of the item (e.g., is it an "easy/hard" item; is one response more socially desirable than another; what is the probability of guessing the item correctly; how well does the item differentiate between individuals possessing differing amounts of ability). The basic idea being that individuals with higher values of the measured latent trait will be more likely to positively respond to an item than

individuals possessing less of that trait while taking into account the characteristics that influence the item's ability to accurately measure that trait.

To estimate this relationship, the probability of an examinee's response profile (i.e., pattern of observed responses) at a given level of theta is estimated. In other words, it is based on conditional probability: the probability that examinees will positively respond to the item given their previous response patterns and estimates of theta.

Since neither theta nor the characteristics of the item are known, they both must be estimated simultaneously through some sort of iterative process. Ad hoc methods are generally used to obtain initial values for the item parameter and ability (i.e., theta) estimates. For instance, bayesian estimation procedures can be used to incorporate any prior information about the ability distribution, such as information about the mean (as with the Expected A Posteriori [EAP] estimation method) or the mode of the posterior distribution, into the calculation of initial theta estimates (Hambleton et al., 1991).

Once the "starting" values are obtained, the likelihood equations for the item parameter estimates are solved by holding the initial ability parameter estimates constant. These resulting item parameter estimates are then used to obtain revised estimates of the ability parameters, which in turn are used to reestimate the item parameters; and so on. This procedure continues until the changes in the ability and item parameter estimates from one iteration to the next is small enough to say the estimates have converged to their "true" values (Hulin et al., 1983).

Besides assuming a nonlinear relationship between the item and the trait being measured, IRT has a number of other advantages over CTT approaches. For instance, its primary focus is at the item level of analysis as opposed to the composite level (e.g., total test score) of analysis, as is the case with CTT. For instance, with a 3-parameter logistic IRT model, three different sources of information -- or item parameters -- concerning an item's functioning in relation to theta, can be used to evaluate differential functioning among the items in a test pool. These three parameters, which are used to define the shape of each item's ICC, are: (a) discrimination, (*a* parameter), which denotes the degree of information provided by an item in relation to theta; (b) difficulty (*b* parameter), which determines the location on theta where the item is most informative or discriminating (e.g., *a* parameter) and is, therefore, the point on theta where the item exhibits its maximum accuracy, or freedom from error; and (c) pseudoguessing (*c* parameter), which is used to compensate for the probability that an observed response is a function of any factor other that rather than a examinee's true standing on theta (e.g., guessing, social desirability). For instance, if an item for the T-F dimension read "In general, are you more apt to watch a story on television that is: (a) dramatic and touching, or (b) informative and convincing?" and had a c parameter of .27 this would indicate that even "thinking" people have a 27% probability of endorsing this item in the feeling direction. An item with a c parameter near zero (e.g., .04) indicates that there is a very small chance that a thinking person would endorse the feeling response as opposed to the thinking response for that item. Item-based statistics have a number of applications such as the designing of tailored tests

for specific purposes (e.g., tests that discriminate well around the cutoff score for high ability examinees), and building item banks for computer adaptive testing (Hambleton et al., 1985, 1991).

An important limitation of CTT is that item and test characteristics are sample-dependent; that is, whether an item or test is difficult, as defined by item difficulty and number-right statistics, is based solely on the ability of the group of examinees in which it was administered . For example, if the same test was administered in two groups -- one of below-average ability and the other of above-average ability - the p (difficulty) values for the same set of items will be much higher in the low ability group than in the high ability group.

Similarly, the scores of examinees are test-dependent; that is, they can only be meaningfully interpreted within the context of the test in which they were assessed. This makes comparison of examinee scores across tests very difficult because the tests are on different scales (Hambleton, 1991). For instance, if two groups of examinees with the same level of ability were given the same test but one test was normed in a sample possessing below-average ability and the other test was normed in a sample possessing above-average ability -- because the test and item characteristics are totally dependent upon the mean level of trait possessed by the norming group -- subsequent examinees taking the test normed on the lower-ability group would appear to have more ability than a group of examinees taking the test normed on the higher-ability group even though the two group's "true" theta level is identical.

Unlike CTT, IRT item and test indices are not sample-dependent. Similarly, examinee scores, or ability estimates, are not test-dependent. According to Hambleton (1991), "ability estimates obtained from different sets of items will be the same (except for measurement error), and item parameter estimates obtained in different groups of examinees will be the same (except for sampling error)" (p. 8). This property of item and ability invariance is due to the joint estimation of ability and item parameter estimates, as described above, because information about each is used in the estimation of the other (e.g., ability parameter estimates are used in the calculation of item parameter estimates and vice versa; Hambleton, 1991). As a result, scores from two different tests can be placed on the same theta metric, items that function differently for various subgroups (i.e., item bias or differential item functioning [DIF]) can be detected, and it is possible to determine whether translations of tests into different languages are comparable.

Another distinction between IRT and CTT is the concept of standard error of measurement (SEM). With CTT, the SEM is a function of both the test's reliability and variance, and is assumed to be the same for all examinees. But test scores for examinees of different levels of abilities contain different amounts of error. For instance, if an examinee gets none of the questions on an ability test correct, then it is assumed that the examinee has a low ability level but how low cannot be known since there is no information pertaining to what the examinee can do. If, however, the examinee gets some items right and some items wrong, then the test score provides more information about

what that examinee can and cannot do and therefore provides more information -- and, accordingly, a more precise measure -- of that examinee's ability level.

To put it another way, with a normal distribution, the most precise measurement (maximum about of information) occurs at the average ability level because, by definition, that is the point where the most data (information) exists about the ability being measured; therefore it is the distribution point least likely to suffer from sampling error. Similarly, as an individual's ability level departs from the average, the test becomes less precise because the standard error also becomes larger with departures from the average. A major advantage with IRT is that the SEM varies as a function of the underlying trait being measured with the lowest SEM at the point at which the test provides its maximum amount of information.

One of the few MBTI studies that adopted an IRT approach, Harvey and Murry (1994), used a 3-parameter logistic IRT model, that the "MBTI items do not contribute equally to the measurement of the dimensions" (p. 8). For example, discrimination (a) parameters for various items ranged from .35 to 2.1 on the E-I scale alone, meaning that some items discriminated along the E-I dimension very well, while others contributed relatively little to its measurement of the underlying trait. In addition, many of the item location/difficulty (b) parameters indicated that many of the items discriminated poorly around the type cutoff scores.

Although the point of maximum discriminating power/measurement precision for each MBTI scale was quite close to the type cutoff scores -- which is desirable, in the

sense of maximizing the likelihood of assigning examinees to the correct dichotomous type categories -- the amount of information provided by each scale at its point of maximum information left much room for improvement. That is, at the point of maximum precision, standard errors ranged from .31 - .34, so that at best the MBTI scales would produce a confidence interval approximately two thirds of a standard deviation in width at the point at which preference scores were dichotomized (i.e., cutoff score with a confidence interval of +/- .34). Based on the findings of this study it is evident that higher measurement precision in the middle region of each MBTI scale is desirable because it would reduce the number of type misclassifications that are caused by measurement error in the vicinity of the cutoff score.

Structural Validity

As discussed earlier, the question of whether the MBTI actually measures Jungian theory has been a very controversial one (see Block & Ozar; 1982; Coan, 1978; Mendelsohn et al., 1982; Murray, 1990; Sipps & Dicaudo; Stricker & Ross, 1964; Weiss et al., 1982). A number of studies have utilized factor analytic techniques to study the structural validity of the fundamental assumption that four personality dimensions underlie the MBTI's construction. Recent exploratory factor analysis conducted by Harvey, Murry, and Stamoulis (in press) yielded "a textbook 4-factor solution exhibiting very clear simple structure and near-perfect consistency with the predicted pattern of loadings" (p.16).

Previous factor analytic examinations resulted in mixed support for the four-factor model. For instance, Sipps, Alexander, and Friedt (1985) conducted a principal components factor analysis yielding a six factor solution, four of which matched the MBTI scales. Comrey (1983) reported a five factor solution which resembled the dimensions of the MBTI, with the T-F scale split into two factors. Thompson and Borrello conducted an exploratory (1986) and later a confirmatory (1989) factor analysis and in both studies they recovered a four-factor structure. Tzeng, Outcalt, Boyer, Ware, and Landis (1984) also reported four "clear simple structures" were found from their factor analytic study.

Although a number of these studies seem to support a four factor model, specific conclusions based on any of these previous studies is problematic, at best, since many of the studies were either methodologically questionable (Comrey, 1983; Sipps, Alexander, & Friedt, 1985; Thompson & Borrello, 1986, 1989), or failed to report their analytic procedures (Tzeng, Outcalt, Boyer, Ware, & Landis, 1984), thereby making assessment of their results impossible (for further discussion, see Harvey, Murry, & Stamoulis, in press).

Based primarily on Harvey et al. (1994) findings and the more general conclusions of the previous factor analytic studies, the validity of the four dimensional model of the MBTI appears to be supported, with the caveat that further research on this issue is still needed. The question then becomes, what do these four factors measure? An inspection of the factor loadings from the Harvey, Murry, and Stamoulis (in press) study coincide with the findings of other studies (Coan, 1978; Comrey, 1883; McCrae & Costa, 1989; Murray, 1990; Sipps & DiCaudo, 1980; Stricker & Ross, 1964) indicating that the

constructs being measured are narrower than those proposed by Jung. Therefore, although the test seems to be measuring four factors as purported by its developers, as previously mentioned, what these factors are measuring in terms of the underlying constructs may be somewhat, if not significantly, different from what has been proposed from a theoretical basis by its authors.

Utility

A last issue is the test's utility. While utility may not be a traditional psychometric issue (Crocker & Algina, 1986), if a test is to be used in industry -- where cost-benefit ratios decide an organization's ability to survive -- then utility is of the utmost importance. Millions of dollars are spent each year to assess employees' personality types and yet much of the research that has been done on the MBTI seems to show that companies may not benefit from using the MBTI (especially for selection and promotion purposes) for two major reasons: (a) it is unreliable -- the MBTI is especially imprecise around the cutoff score and, as a result, individuals have about a 50% chance of being assigned to a different personality type the next time they take the test (even if the next administration occurs within a 5-week time frame); and finally, (b) the constructs being measured by the test may only represent limited aspects of what the test developers propose they are measuring (i.e., the test does not adequately operationalize Jung's type theory) and, therefore, interpretations based on strictly Jungian type theory may be inappropriate.

Purpose of the Present Study

Because much of the available MBTI research strongly suggests the need to "either improve the MBTI or offer an alternative measurement techniques allowing for greater sensitivity" (Cowan, 1989, p. 461), we assessed the degree to which we could improve its measurement properties by writing additional items for each of the four primary MBTI scales. As earlier research has shown, this need for additional test information and reduced SEMs is most critical in the middle regions of each scale (i.e., near the type cutoff score). To improve its measurement properties new items were written for each of the MBTI scales as a means to increase the amount of test information provided by each MBTI scale. Two hundred new items -- 50 per scale -- were written. This was accomplished by first examining those items that seemed most definitive of the four MBTI scales, based on prior psychometric studies and factor analyses. The most definitive items were deemed to be those that demonstrated the highest (a) IRT a (discrimination) parameters (values of .90 and up were considered high) from the Harvey and Murry (1994) study, and (b) factor loadings from the Harvey et. al. (in press) study.

Insert Table 1 about here

New items were then written to parallel the general substantive content of these "high-performing" items. These new items were retained if they (a) paralleled the constructs measured by the high-performing original MBTI items, and (b) substantially increased the

amount of information provided by the MBTI, especially around zero on the theta scale (e.g.,. the point at which the continuous scores of each dimension are dichotomized to assign categorical types).

To evaluate the new items, two strategies were followed. First, item-total correlations (e.g., between each new item, and the composite formed from the high-performing items from the original item pool) were examined to screen out new items that appeared to be functioning poorly. Items that passed the item-total criterion were then analyzed via exploratory factor analysis conducted in each item pool (e.g., containing both old and new items) to assess the degree to which the new items measured the same dimensions as the original items and the degree to which the IRT assumption of unidimensionality was met. The unidimensionality assumption of IRT only requires the existence of a dominant factor to account for test performance (Hambleton et al., 1991) Because the assumption of a truly unidimensional latent space is generally too strong for most data sets from applied settings (Hulin et al., 1983), a number of studies have been conducted to demonstrate that IRT models are robust to moderate violations of unidimensionality (Drasgow & Parson, 1983; Harrison, 1986; Lim & Drasgow, 1992).

Second, the amount of information provided by the new items versus the original items was quantified using IRT techniques to produce test information functions (TIFs). TIFs are graphical representations of the amount of information provided by all levels of theta, and are roughly comparable to the notion of reliability in classical test theory (e.g., the more information provided by a test at a given point on the theta continuum, the more

precise -- and therefore reliable -- it should be in terms of estimating the underlying trait). By comparing TIFs for different item pools (e.g., original items only, new items only, and old plus new items), the degree to each MBTI scale was improved via the addition of new items could be quantified.

Method

Subjects and Instrument

A sample of 583 undergraduates at Virginia Polytechnic Institute and State University participated voluntarily in this study in exchange for course credit. Subjects were administered two hundred new items that were randomly intermixed along with 94 of the 95 marker items of the MBTI Form F (item 68 was dropped during prior IRT item analysis because it allowed more than one response alternative to be marked; see Harvey & Murry, 1994). Items were given a value of 1 if the person's response was in the I, N, F, or P direction, and 0 if the response was in the E, S, T or J direction.

Analyses

For the item-total correlations, composites of high-performing items were formed, consisting of the 10 best items from each scale of the original MBTI as determined during previous analyses (i.e., based on high factor loadings and high a parameters). Item-total correlations were then produced for each 10-item composite and its corresponding 50 new items. For the exploratory factor analysis, four item pools, consisting of the 10 highest-loading items from original scale plus the new items retained for that scale, were analyzed separately. To confirm the existence of one primary factor for each scale, items

were analyzed to ensure they had strong factor loadings on the first unrotated principle axis; eigenvalue discontinuities (i.e., the "scree test") were also examined to confirm that the item pool defined a single primary dimension. The 3-parameter logistic IRT model was used to calibrate items (items from the original scale, and the selected new items that passed the above tests, were calibrated jointly). Item parameters were estimated using BILOG (Mislevy & Bock, 1990); the EAP method with the biweight option was used to estimate theta scores. This analysis yielded the a, b, and c parameters needed to generated the TIFs for scales formed by various combinations of old and new items.

Results and Discussion

Dimensionality

New items that exhibited less than approximately a .40 correlation with their high-performing item composite were dropped from further analysis. Using this criteria, 64 new items were dropped: 15 from the E-I scale, 12 from the S-N scale, 15 from the T-F scale, and 20 from the J-P scale. Figure 1 indicates the frequency distribution of the item-total correlations prior to dropping the 64 new items (see Figures 1-4).

Insert Figures 1-4 about here

The remaining new items and the original items were then used in the exploratory factor analysis. The percentages of common variance accounted for by the first unrotated principle axis for each item pool were 70%, 68%, 79% and 78% for the E-I, S-N, T-F,

and J-P scales, respectively. Given the large size of these items pools, and the relatively large size of the first versus the second eigenvalues, these percentages were viewed as being consistent with the conclusion that each item pool primarily assessed a single latent trait and was sufficiently unidimensional for IRT analyses.

Test Information Functions

The item parameter estimates used to generate the TIFs showed that a number of the new items were more discriminating than the original items (i.e., had higher a parameters). Figures 5-52 show the frequency distributions of the a, b, and c parameters for four different item pools within each scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting all of the new items analyzed using IRT techniques, (c) a third item pool consisting of new items in equal number to those contained in the original scale, and (d) a fourth item pool consisting of the original items plus all of the new items. These distributions indicate that the a parameters of the new items were considerably and consistently higher than those of the original items. The b and c parameters across both sets of items were generally comparable (see Figures 5-52).

Insert Figures 5-52 about here

Likewise, TIFs were generated for four different item pools within each scale: (a) one item pool consisting of only the original items (22 for the E-I scale, 23 for the T-F and J-P scales, and 26 for the S-N scale), (b) a second item pool consisting all of the new

items analyzed using IRT techniques (35 for the E-I scale, 38 for the S-N scale, 35 for the T-F scale, and 30 for the S-N scale), (c) a third item pool consisting of new items in equal number to those contained in the original scale (i.e., 22 for the E-I scale) to allow a cross-scale comparison using an equal number of items (i.e., if the numbers of new versus old items for each MBTI scale differed, any superiority of the new versus the old scale might be due simply to an increased number of items in the new scale), and (d) a fourth item pool consisting of the original items plus all of the new items (see Figures 53-56).

Insert Figures 53-56 about here

Inspection of the TIFs indicated a substantial gain in information provided by the new items: across all four scales, the item pool containing only the original items provided the least amount of information, the new item pool that contained as many new items as there were old items provided substantially more information than the old items alone, and the item pool containing both the new and old items provided the most information. The TIFs provided their maximum amount of information around zero theta units, which is consistent with the goal of maximizing the amount of information regarding the latent trait around the cutoff score that would be used to form categorical type assignments.

In terms of the SEMs for each MBTI scale, the results in Figures 53-56 indicate that the addition of new items produced a substantial decrease in standard errors of estimating theta (computed as 1/sqrt of theta), especially in the critical area around the

type cutoff scores. For example, the TIFs for the EI scale (Figure 53) indicated that for the original item pool, the TIF peaked at about 12 information units (e.g., an SEM of approximately .29 z units); in contrast, the scales formed using an equal number of new items (information = 36, SEM = .17) and the entire new item pool (information = 42, SEM = .15), and the combined old and new pools (information = 53, SEM = .14) offered an appreciably higher level of measurement precision at this point. Indeed, at their peaks, the SEM for the full new-item pools and the old-plus-new pools were approximately half the size of the SEM for the original item pool, reflecting a substantial improvement in measurement precision due to the new items. Inspection of the SEMs corresponding to the TIFs contained in figures 53-56 further supported this contention (see Figures 57-60).

Insert Figures 57-60 about here

Conclusions

Our results clearly indicate that: (a) new item pools can indeed be written to parallel the item content exhibited by the four main MBTI scales, and, of greater importance, (b) these new items dramatically increase the amount of test information (and decrease the SEMs) relative to the original item pools. The increase in test information functions would allow significantly more precise judgments to be made regarding type classifications, especially for individuals whose theta scores lie close to the cutoff score on each dimension. It remains to be seen, through future research, whether studies

conducted using these new item pools will realize corresponding improvements in test-retest reliability (especially for the categorical type assignments) due to this increased measurement precision.

In addition, the existence of new item pools to assess the four main MBTI constructs -- and their clear superiority vis a vis test information and SEM functions over the basic Form F item pools, even when the number of items is identical between scales formed using the new versus the old item pools -- will be important in allowing the development of a computerized adaptive test (CAT) to assess the MBTI's underlying four constructs. Past research (e.g., Harvey et. al., 1994) has suggested that a CAT version of the MBTI could be developed that would (a) provide a level of measurement precision that is comparable to the full-length instrument, and (b) significantly reduce administration time (i.e., CAT methods customize the presentation of items to each examinee based on their previous responses, which can result in far fewer items being presented than in a paper-and-pencil versions). That is, many of the forms of the MBTI are quite long (e.g., Form F contains 166 items, Form J contains 281 items), and the MBTI lacks a lie scales or a similar means for detecting aberrant response profiles; a CAT implementation could address both of these needs, yet still produce measurement precision comparable to the full-length instrument.

In addition, a CAT version of the MBTI could be used to produce IRT appropriateness indices to quantify the degree to which a response profiles are free from internal inconsistencies or other aberrant response patterns (e.g., Hulin et al., 1983). These

indices could be used to detect faking, an important testing issue especially in light of Furnham's (1990) observation that the particular personality tests used most frequently in industry seem to be more susceptible than other measures to dissimulation. To underscore the necessity of such faking indices further, it has been demonstrated that examinees can fake their scores on the MBTI (Furnham, 1990), even to the extent of matching certain response profiles (e.g., rather than the usual "fake good/fake bad" paradigm used in faking studies) and that some examinees are significantly better at faking than others (Snell & Harvey, 1994). Although some argue that the majority of actual job applicants appear to refrain from such response distortion (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), whenever a personality test is used in a situation where the examinee might be motivated to fake test scores and the cost to the organization for making an incorrect decision based on these test scores is significant (e.g., for selection and promotion decisions), the possibility of faking must be taken into consideration.

In terms of the MBTI's relation to the Big Five, this study also demonstrated that it is possible to successfully write new items using the methods utilized in this study and, moreover, that it is possible to create additional items to broaden the content domain of the existing five scales (especially the "new" Neuroticism scale), as suggested by Harvey et al. (1994), so that the MBTI would measure the Big Five factors as comprehensively as other existing instruments developed specifically for that purpose. With the growing acceptance of the five-factor structure of personality (Cortina et al., 1992; Digman, 1990;

Hogan, 1991; McCrae & Costa, 1987; Schmit & Ryan, 1993), there are a number of reasons for adopting a Big Five interpretation of the MBTI.

For instance, McCrae and Costa (1989) contend that a five-factor interpretation would provide a more "meaningful framework in which to organize and interpret the many correlates -- a set of standard dimensions by reference to which the MBTI indices can be understood" (p. 23). Although the Big Five model is only descriptive in nature in that it does not explain why (e.g., the origins or the underlying mechanisms) there are five dimensions of personality, McCrae and Costa (1989) -- commenting on the number of studies in which the findings were not explained by, and often contrary to, Jungian theory -- argue that Jungian psychology, as measured by the MBTI, cannot provide a better theoretical basis for differences in personality and at least the Big Five has strong empirical support. Within a Big Five framework, a number of the anomalous correlations reported in the manual could be better understood. For instance, the MBTI T-F dimension is positively related to affiliation, blameavoidance, nurturance, and succorance and is negatively related to aggression, counteraction, and dominance. These correlations would not be predicted from Jung's definition of the T-F dimension but from a five-factor perspective, they are understandable if viewed in relation to the studies showing that the T-F dimension of the MBTI corresponds to the Agreeableness dimension of the Big Five and that these "subscales" represent an interpersonal aspect of the Agreeableness dimension.

From a more applied perspective, if the MBTI were being used primarily as a predictive instrument (e.g., selection), using the continuous scores from the MBTI based on a Big Five interpretation would allow MBTI users to take advantage of the growing empirical and meta-analytic evidence supporting the Big Five's validity in predicting various aspects of job performance (Barrick & Mount, 1991; Cortina, Doherty, Schmitt, Kaufman, & Smith, 1992; Mount, Barrick, & Strauss, 1994; Schmitt & Ryan, 1993; Tett, Jackson, & Rothstein, 1991). Using these continuous scores would further allow MBTI use all of the information provided by the scales (since, as previously mentioned, some information loss is to be expected when they are dichotomized) and provide a useful alternative to the controversial categorical type-based scoring system currently in use (Harvey & Murry, 1994; Harvey et al., 1994, Murray, 1990). That is, although Harvey & Murry (1994) demonstrated that an IRT scoring system, rather than the traditional preference scoring system, showed a bimodal distribution scores (which would support the existence of discrete personality types), they cautioned that there was still some measurement uncertainty for those examinees who occupied the minimum density point between the two modes (i.e., the discontinuities at the midpoints of each scale was not sharp enough to indicate zero SEMs). Especially for those occupying this "zone of uncertainty", a continuous scoring method would still provide more information about their level of theta than the traditional dichotomous scoring method.

In addition, for those employers utilizing personality measures for selection purposes, using the MBTI to derive a five-factor score, as opposed to using other

five-factor instrument, may have another important advantage. The issue of an applicant's right to privacy has recently become a legal issue surrounding the use of personality instruments for selection purposes (e.g., Saroka v. Dayton Hudson). Because the item content of the MBTI is less intrusive than many of the other personality, especially those asking questions about sensitive topics (e.g., questions about sexual orientation, religious beliefs, and bodily functions) such as the Minnesota Multiphasic Personality Inventory (MMPI) or California Psychological Inventory (CPI), questions concerning invasion-of-privacy issues are probably less likely to arise with the use of the MBTI. Harvey et al. (1994) contends that "using the MBTI's item pool to estimate scores on the Big Five dimensions might offer significant advantages in the area of legal defensibility" (p. 12). For these reasons, using the MBTI to derive a five-factor score may be less risky from a legal standpoint than using another instrument to derive scores for the Big Five.

Although these are all strong reasons for adopting a five-factor interpretation of the MBTI, completely abandoning the traditional type-based interpretation, as suggested by McCrae and Costa (1989), may not be warranted, nor advisable, in some situations. As pointed out by Harvey et al. (1994), the traditional method of interpreting MBTI scores as proposed by the test's authors, although perhaps not a veridical translation of Jungian theory, may have some advantages over a Big Five interpretation in situations where the instrument is being used to increase understanding among employees. Under these circumstances, personality information needs to be conveyed in a manner that is "positive"

and "affirming" for each participant. Similarly, being told one is neurotic from a Big Five perspective could hardly be construed as being positive and affirming.

In addition, conveying this information through the use of types makes the information easier to understand for the participants, especially since many people are already familiar with the MBTI types. The question of whether sixteen qualitatively distinct types exist is a controversial one that many argue has not been supported by empirical evidence (Hicks, 1984; Murray, 1990; McCrae & Costa, 1989; Stricker & Ross, 1964; Wiggins, 1989). However, Snell and Harvey (1994), found that temperament types (e.g., the presence of either the SJ, SP, NT, or NF combination in a person's four-letter type) were significant predictors of peoples' ability to fake opposite preference scores on the MBTI. In addition, it is important to note that previous research has generally assumed that the presence of a significant statistical interactions is necessary and sufficient to demonstrate the existence of distinct types (Hicks, 1985; Mendelsohn, Weiss, & Feimer, 1982; Weiss, Mendelsohn, & Feimer, 1982). However, as pointed out by Pervin and Lewis (1978), the absence of such findings does not preclude the existence of an interactive relationship among the variables because traditional approaches to detecting interactions assume that the relationship between the variables is unidirectional and static even though many of the relationships between variables found in the social sciences tend to be bidirectional and dynamic.

Finally, an issue that cannot be addressed by the present research is the issue of type indeterminancy -- that some individuals do not have a true preference for either pole

(e.g., E vs. I) for one or more of the MBTI dimensions. Studies have indicated that examinees who have higher preference scores (e.g., 42 for the E pole) are less likely to change type upon the next test administration (Carskadon, 1979; Myers & McCaulley, 1985). In fact, in the research report earlier concerning the percentage to test-retest agreement in each category, which generally ranged from some lower percentage (e.g., 54%) to 100% agreement, those who had 100% agreement in each category were those examinees whose initial preference scores were high (i.e., 31 and above; Myers & McCaulley, 1985). Likewise, those with low initial preference scores (i.e., 0 -15) had the lowest percentage of test-retest agreement and those with moderate preference scores (i.e., 16-29) occupied an intermediate position relative to those with high and low preference scores across all four categories (i.e., E-I, S-N, T-F, and J-P).

It is important to note that this type of research cannot determine whether changing one's type is due to not having a relative preference, as in the case of type indeterminancy, or the result of careless test taking or measurement error around the cutoff score. One possible means of studying the issue of type indeterminancy is through the use of self-reports and IRT appropriateness indices. Appropriateness indices provide empirical evidence as to whether or not a particular test score is a valid measure for individuals by examining whether they responded to items in an internally consistent manner (e.g., cheating may be inferred from a low ability examinee answering a number of very "high ability" items correctly). Specifically, appropriateness measures examine the

probability of a response pattern based on the difficulty (b) parameters of the answered items.

By using these indices in conjunction with self-reports, the response pattern of those with a "bi-preference" (e.g., respond equally positively to items with high b parameters for both poles of a dimension) versus individuals who do not have a strong preference either way (e.g., respond equally positively to items with low b parameters for both poles of a dimension) can be studied. Adaptive testing would be the most efficient means of accomplishing such research because the difficulty of the test -- as indicated by the location on theta where the test offers its maximum amount of information -- could be tailored to the ability of the examinee and, therefore provide maximum information at the point on the theta continuum where the respondent's ability level is thought to lie.

In summary, regardless of how the MBTI is being used (e.g., as a Big Five instrument or not) or for what purposes it is used (e.g., for prediction or understanding of behavior), this research has shown that a more reliable instrument can be constructed. Because a reliable instrument is a necessary prerequisite to the establishment of the instrument's validity, the question of whether the MBTI can predict different aspects of an individual's personality can now be effectively studied. Whether studied via CAT, a five-factor scoring system, or traditional means, more empirical support for its validity, in terms of knowing what aspects of an individual's behavior the MBTI does and does not predict, would increase its usefulness for organizations and individuals alike.

References

Agor, W. H. (1988). Finding and developing intuitive managers. Training and Development Journal, 42, 68-70.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance. Personnel Psychology, 44, 1-26.

Block, J., & Ozer, D. J. (1982). Type types of psychologists: Remarks on the Mendelsohn, Weiss, and Feimer contribution. Journal of Personality and Social Psychology, 42, 1171-1181.

Briggs, K. C., & Myers, I. B. (1976). Myers-Briggs Type Indicator: Form F. Palo Alto: Consulting Psyhologists Presss.

Bushe, G. R. & Gibbs, B. W. (1990). Predicting organizational development consulting competence from the Myers-Briggs Type Indicator and stage of ego involvement. Journal of Applied Behavioral Science, 26, 357-378.

Carlson, J. (1985). Recent assessments of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 49(4), 356-365.

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 41, 461-473.

Carskadon, T. G. (1977). Test-retest reliabilities of continuous scores on the Myers-Briggs Type Indicator. Psychological Reports, 41, 1011-1012.

Coan, R. W. (1978). Review of the Myers-Briggs Type Indicator. In O. K. Buros (Ed.), The Eighth Mental Measurements Yearbook. Highland Park, NJ: Gryphon Press. pp. 970-975.

Coe, C. K. (1992). The MBTI: Potential uses and misuses in personnel administration. Public Personnel Management, 21(4), 511-523.

Comrey, A. L. (1983). An evaluation of the Myers-Briggs Type Indicator. Academic Psychology Bulletin, 103, 115-129.

Cortina J. M., Doherty, M. L., Schmitt, N, Kaufman, G., & Smith, R. G. (1992). The "Big Five" personality factors in the IPI and MMPI: Predictors of police performance. Personnel Psychology, 45, 119-140.

Costa, P. T., & McCrae, R. R., (1985). The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources.

Cowan, D. A. (1989). An alternative to the dichotomous interpretation of Jung's psychological functions: Developing more sensitive measurement technology. Journal of Personality Assessment, 53, 459-471.

Crocker, L. & Algina, J. (1986). Introduction to classical & modern test theory. Orlando, FL: Holt, Rinehart, and Winston, Inc.

Davis, D. L., Grove, S. J., & Knowles, P. A. (1990). An experimental application of personality types as an analogue for decision-making style. Psychological Reports, 66, 167-176.

Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. Personnel Psychology, 42, 25-35.

DeVito, A. J. (1985). Review of the Myers-Briggs Type Indicator. In J.V. Mitchell, Jr. (Ed.), Ninth Mental Measurements Yearbook, Lincoln, NE: University of Nebraska Press. Vol. 2, pp. 1030-1032.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. Annual Review of Psychology, 41, 417-440.

Drasgow, F., & Hulin, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. Applied Psychological Measurement, 7, 189-199.

Drasgow, F., & Lissak, R. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. Journal of Applied Psychology, 68, 363-373.

Furnham, A. (1990). The fakability of the 16PF, Myers-Briggs and FIRO-B personality measures. Personality and Individual Differences, 11, 711-716.

Garden, A. (1989). Organizational size as a variable in type analysis and employee turnover. Journal of Psychological Type, 17, 3-13.

Gauld, V., & Sink, D. (1985). The MBTI as a diagnostic tool in organization development interventions. Journal of Psychological Type, 9, 24-29.

Gough, H. G. (1976). Studying creativity by means of word association tests. Journal of Applied Psychology, 61, 348-353.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.

Hall, C. S. and Nordby, V. J. (1973). A primer of Jungian Psychology. New York: New American Library.

Hambleton, R. K., & Swaminathan, H. (1985). Item Response Theory: Principles and applications. Boston: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications, Inc.

Harrison, D. A. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. Journal of Educational Statisiics, 11, 91-115.

Hartzler, G. J. & Hartzler, M. T. (1982). Management uses of the Myers-Briggs Type Indicator. Research in Psychological Type, 5, 20-29.

Harvey, R. J., & Murry, W. D. (1994). Scoring the Myers-Briggs Type Indicator: Empirical comparison of preference versus latent-trait methods. Journal of Personality Assessment, 62, 116-129.

Harvey, R. J., Murry, W. D., & Markham, S. E. (1994). Evaluation of three short-form versions of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 63(1), 181-184.

Harvey, R. J., Murry, W. D., & Stamoulis, D. (in press). Unresolved issues in the dimensionality of the Myers-Briggs Type Indicator. Educational and Psychological Measurement.

Harvey, R. J., Murry, W. D., & Markham, S. E. (1994). A "Big Five" scoring system for the Myers-Briggs Type Indicator. Paper submitted to 1995 SIOP conference: Orlando, FL.

Hicks, L. E. (1984). Conceptual and empirical analysis of some assumptions of an explicitly typological theory. Journal of Personality and Social Psychology, 46, 1118-1131.

Hicks, L. E. (1985). Dichotomies and typologies: Summary and implications. Journal of Psychological Type, 10, 11-13.

Hogan, R. (1991). Personality and personality measurement. In M. Dunnette and L. Hough (Eds.), Handbook of industrial and organizational psychology (2ed.). Palo Alto: Consulting Psychologists Press.

Hogan, J. & Hogan, R. (1988). How to measure employee reliability. In Employee testing: The complete resource guide. Washington, D.C.: Bureau of National Affairs, Inc.

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

Hoy, F. & Hellriegel, D. (1982). The Kilmann and Herdon model of organizational effectiveness criteria for small business managers. Academy of Management Journal, 25, 308-322.

Hulin, C., Drasgow, F., & Parsons, C. (1983). Item response theory: Application to psychological measurement. Homewood, IL: Dow Jones-Irwin.

Johnson, D. A., & Saunders, D. R. (1990). Confirmatory factor analysis of the Myers-Briggs Type Indicator -- expanded analysis report. Educational and Psychological Measurement, 43, 116-135.

Jung, C. (1923). Psychological types. New York: Harcourt, Brace & Co.

Jung, C. (1971). Psychological types. Princeton, NJ: Princeton University Press.

Kirton, M. J. (1976). Adaptors and innovators. Journal of Applied Psychology, 61, 622-629.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1992). Predictors of individual differences in mail-coding skills and their variation with ability levels. Journal of Applied Psychology, 77(4), 406-418.

McCarley, N., & Carskadon, T. G. (1983). Test-retest reliablities of scales and subscales of the Myers-Briggs Type Inventory and of criteria for clinical interpretive hypotheses involving them. Research in psychological type, 6, 24-36.

McCaulley, M. H. (1990). The Myers-Briggs Type Indicator and Leadership. In K.E. Clark & M.B. Clark (Eds.), Measures of Leadership, (pp. 381-418). Greensboro, NC: Center for Creative Leadership.

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, 347-367.

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. Journal of Personality, 57, 17-40.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel Psychology, 43, 335-354.

Mendelsohn, G. A., Weiss, D. S., & Feimer, N. R. (1982). Conceptual and empirical analysis of the typological implications of patterns of socialization and feminity. Journal of Personality and Social Psychology, 42, 1157-1170.

Mislevy, R. J., & Bock, R. D. (1990). BILOG 3: Item analysis and test scoring with binary logisitc methods. Mooresville, IN: Scientific Software.

Moore, T. (1987). Personality tests are back. Fortune. March 30, 74-82.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. Journal of Applied Psychology, 79, 272-280.

Murray, J. B. (1990). Review of the research on the Myers-Briggs Type Indicator. Perceptual and Motor Skills, 70(3), 1187-1202.

Myers, I. B. (1962). The Myers-Briggs Type Indicator manual. Princeton, NJ: Educational Testing Service.

Myers, I. B., & McCaulley, M. H. (1985). Manual: A guide to the developemt of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press, Inc.

Myers, I.B. and Myers P.B. (1980). Gifts Differing. Palo Alto, CA: Consulting Psychologists Press.

Palmiere, L. (1972). Intro-extraversion as an organizing principle in fantasy production. Journal of Analytical Psychology, 17, 116-135.

Pattison, P. & Varca, P. E. (1991). The demise of the disparate impact theory. American Business Law Journal, 29, 413-450.

Pervin, L. A. & Lewis, M. (1978). Perspectives in interactional psychology. New York: Plenum Press.

Poilitt, I. (1982). Managing differences in industry. Research in Psychological Type, 5, 4-19.

Ramaprasad, A., & Mitroff, I. I. (1984). On formulating strategic problems. Academy of Management Review, 9, 597-605.

Sample, J. A., & Hoffman, J. L. (1986). The MBTI as a management and organizational tool. Journal of Psychological Type, 11, 47-50.

Saunders, D. (1986). Type differentiation indicator manual: A scoring system for Form J of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Schmitt, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structures in applicant and nonapplicant populations. Journal of Applied Psychology, 78, 966-974.

Sipps, G. J., Alexander, R. A., & Friedt, L. (1985). Item analysis of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 45, 789-796.

Steele, R. S., & Kelly, T. J. (1976). Eysenck personality questionnaire and Jungian Myers-Briggs Type Indicator of extraversion-introversion. Journal of Consulting and Clinical Psychology, 44, 690-691.

Stricker, L. J., & Ross, J. (1964). Some correlates of a Jungian personality inventory. Psychological Reports, 14, 623-643.

Suplee, C. (1991, September 25). Performance-enhancement techniques found flawed or ineffective in study. The Washington Post, p.A3.

Tett, R. P., Jackson, D. N., Rothstein, M. & Reddon, J. R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). Personnel Psychology, 47, 157-172.

Thompson, B., Borrello, G. M. (1986). Construct validity of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 46, 745-752.

Thompson, B., Borrello, G. M. (1989, January). A confirmatory factor analysis of data from the Myers-Briggs Type Indicator. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston.

Tzeng, O. C. S., Ware, R., & Bharadwaj, N. (1991). Comparison between continuous bipolar and unipolar ratings of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 51, 681-690.

Tzeng, O. C. S., Ware, R., & Chen, J. (1989). Measurement and utility of continuous unipolar ratings for the Myers-Briggs Type Indicator. Journal of Personality Assessment, 53, 727-738.

Tzeng, O. C. S., Outcalt, D., Boyer, S. L., Ware, R., & Landis, D. (1984). Item validity of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 48, 255-256.

Tzeng, O. C. S. (1983). A comparative evaluation of four response formats in personality ratings. Educational and Psychological Measurement, 43, 935-950.

Weiss, D. S., Mendelsohn, G. A., & Feimer, N. R. (1982). Reply to the comments of Block and Ozer. Journal of Personality and Social Psychology, 42, 1182-1189.

Wiggins, J. S. (1989). Review of the Myers-Briggs Type Indicator. In J. C. Conoley & J. J. Kramer (Eds.) Tenth mental measurement yearbook. Lincoln, NE: University of Nebraska Press.

Zemke, R. (1992). Second thoughts about the MBTI. Training, 29(4), 43-48.

Table 1

Harris-Kaiser p=.5 Rotated 4-Factor MBTI Loadings for Phi versus Tetrachoric Correlations (Harvey, Murry, & Stamoulis, in press) and 3-Parameter Logistic IRT Model Item Parameter Estimates using EAP Method with Biweight Option for Theta Score Estimates (Harvey & Murry, 1994)

MBTI Item	Tet	Phi	Discrim *a*	Location *b*	Guess *c*
F114 feeling>thinking	.81*	.67*	1.491295	-0.291725	0.031555
F72 warm-hearted>firm-minded	.77*	.51*	1.299103	-1.179181	0.029000
F79 sympathize>analyze	.77*	.58*	1.023051	0.287564	0.006746
F86 touching> convincing	.74*	.61*	1.191666	-0.264045	0.016323
F103 compassion> foresight	.74*	.60*	1.150501	-0.459968	0.048428
F111 gemtle>firm	.73*	.56*	1.125053	-0.824294	0.017268
F26 value sentiment>logic	.70*	.56*	1.141986	0.002253	0.044575
F89 soft>hard	.66*	.48*	0.830140	-1.024520	0.035493
F29 feeling>reasonable	.63*	.50*	0.835846	-0.646642	0.016301
F4 feelings> rights	.60*	.40*	0.707746	-1.595154	0.033476
F105 mercy>justice	.57*	.40*	0.633960	0.748705	0.013646
F100devoted>determined	.56*	.45*	0.637664	-0.297828	0.058102
F81 blessings>benefits	.56*	.42*	0.599880	0.198666	0.010831
F158 unsympathetic>unreasonable	.55*	.38*	0.698081	0.841457	0.055632
F154 heart>head	.53*	.43*	0.730074	-0.063729	0.016847
F133 enough warmth	.50*	.32*	0.527697	-1.827609	0.027839
F91 forgive>tolerate	.50*	.35*	0.624227	1.303494	0.027516
F120 peacemaker>judge	.47*	.32*	0.597856	-1.538030	0.025030
F108 trustful>wary	.46*	.28*	0.478422	-2.405738	0.031551
F84 uncritical>critical	.46*	.36*	0.481874	-0.188165	0.043296
F147 kind>fair	.46*	.26*	0.475283	2.034060	0.014306
F93 who>what	.45*	.34*	0.507547	-1.025405	0.028028
F122 agree>discuss	.32*	.22	0.246095	2.699956	0.057391
N107 create>make	.72*	.54*	1.014811	-1.009148	0.036673
N128 theory>fact	.71*	.56*	1.000738	0.260954	0.024838
N73 imaginative>matter-of-fact	.69*	.50*	1.045406	-1.071940	0.043330
N104 abstract>concrete	.68*	.55*	1.180407	0.147412	0.014828
N145 ingenious>practical	.67*	.55*	1.041940	-0.074221	0.027021
N88 concept>statement	.67*	.52*	0.835322	-0.623942	0.035468
N102 ideas>facts	.63*	.51*	0.941842	-0.582919	0.017013
N76 theory>certainty	.61*	.44*	0.819643	0.874583	0.029365
N2 imaginative>realistic	.59*	.47*	0.938913	0.311736	0.016434
N70 vision>common sense	.59*	.47*	0.701054	0.091659	0.060010
N78 invent>build	.59*	.48*	0.810889	-0.419867	0.013877
N90 design>production	.56*	.45*	0.727612	-0.797245	0.022150
N17 odd/original writing	.54*	.44*	0.726285	0.062749	0.071292

Table 1 (continued)

N64 analyzing problems	.54*	.40*	0.371087	-0.699220	0.073542
N11 invent ways of doing	.54*	.43*	0.611157	-0.147183	0.070889
N98 fascinating>sensible	.54*	.43*	0.765728	-0.816430	0.039723
N119 figuraive>literal	.54*	.44*	0.767314	-0.064686	0.025472
N140 see possibilities>facts	.52*	.33*	0.419053	-2.137094	0.044498
N117 symbol>sign	.50*	.40*	0.616326	-0.573407	0.048981
N149 ideas>feet on ground	.49*	.41*	0.729859	-0.221138	0.040555
N165 orginal>conventional	.49*	.40*	0.713979	-0.383361	0.049616
N37 original>conventional people	.47*	.36*	0.604449	-0.753184	0.093891
N121 change>accept	.46*	.37*	0.455442	-0.094705	0.046811
N53 annoyed if not like theory	.44*	.34*	0.380572	0.607546	0.108211
I25 strangers take effort	.73*	.55*	0.986179	0.700638	0.003794
I33 hard to know	.73*	.58*	1.031848	0.644763	0.006408
I148 get intro>intro others	.67*	.55*	0.850353	0.012023	0.009983
I126 say only to some	.67*	.55*	0.942415	-0.146655	0.006326
I19 bored at parties	.67*	.47*	0.808052	-0.765585	0.013904
I6 talk to known people	.65*	.52*	0.821998	-0.099185	0.014843
I87 reserved>talkative	.65*	.52*	0.908471	0.400700	0.006110
I134 help others have fun	.65*	.52*	0.813448	-0.104074	0.012920
I138 tell interests after time	.60*	.47*	0.747614	-0.287751	0.099510
I106 calm>lively	.59*	.46*	0.719680	0.655401	0.009646
I92 quiet>hearty	.55*	.43*	0.728084	0.617169	0.013706
I95 write>speak	.54*	.42*	0.501059	0.793775	0.013133
I15 keep feelings to self	.54*	.44*	0.799939	-0.059475	0.103932
I66 less enthusiastic	.51*	.40*	0.643768	0.808380	0.005918
I77 theater>party	.49*	.36*	0.464785	1.130596	0.012034
I160 tell feelings friends	.47*	.37*	0.531753	0.903460	0.007566
I41 last to hear	.42*	.33*	0.427931	0.487387	0.011207
I129 not interested in fashion	.37*	.25*	0.360154	-1.491104	0.024331
I47 embarrassed think days later	.30	.23	0.356067	1.266235	0.018760
P85 unplanned>schedule	.77*	.62*	1.878912	0.337362	0.008402
P132 details as go along	.75*	.58*	0.913299	0.671864	0.024040
P60 list not appealing	.69*	.52*	0.932037	0.910054	0.052606
P151 like planned>unpleasant	.68*	.50*	1.003970	1.012953	0.045943
P55 just going>planning	.67*	.55*	0.934070	0.082541	0.022408
P27 prefer free to do fun	.67*	.52*	0.872429	-0.340111	0.045467
P49 daily routine painful	.66*	.43*	1.377629	1.345775	0.040169
P1 schedule cramps me	.66*	.46*	1.282157	1.050512	0.028599
P13 prefer do last minute	.61*	.49*	0.815782	0.277006	0.077414
P35 new projects w/o plan	.60*	.48*	0.620091	0.311732	0.062236

Table 1 (continued)

P74 spontaneous>systematic	.56*	.41*	1.204612	-0.571610	0.021690
P118 casaul>systematic	.55*	.43*	0.967937	-0.439921	0.015737
P94 impulse>decision	.55*	.44*	0.852018	-0.032540	0.016772
P142 speed at end>start early	.53*	.43*	0.642483	0.317861	0.086650
P9 can't tell well about Sat.	.53*	.41*	0.667904	0.382017	0.060451
P42 best at dealing w/unexpect	.51*	.41*	1.182585	0.485342	0.172805
P109 easy-going>orderly	.50*	.34*	0.837920	-1.058589	0.031717
P153 often forget small things	.48*	.35*	0.453622	1.630483	0.034864
P97 leisurely>punctual	.45*	.36*	0.739145	0.200579	0.028998
P113 quick>careful	.43*	.32*	0.430963	1.275948	0.035071
P124 routine pts of day=boring	.38*	.29*	0.539671	0.663294	0.166698
P20 hard to adapt to routine	.36*	.28*	0.780850	1.044865	0.129620
P99 changing>permanent	.34*	.23	0.561896	-0.789507	0.073348

Note. Factor loading entries are the rotated primary factor pattern loadings computed from the matrix of tetrachoric correlations and the matrix of phi correlations; values greater than the root-mean-square loading are marked with an asterisk.

Table 2

Item Parameter Estimates for the New E-I Items

Item Number	Discrimination (*a*) parameters	Location (*b*) parameters	Guessing (*c*) parameters
dL3	0.940840	0.127666	0.056474
dL6	0.930680	0.564136	0.082211
dL15	1.216569	0.097636	0.012483
dL19	2.241320	0.650987	0.002621
dL22	1.240151	0.380080	0.007993
dL25	0.743395	0.675592	0.004147
dL31	0.982783	0.404137	0.011156
dL39	0.894167	0.474708	0.014948
dL42	1.837676	0.449312	0.007595
dL45	1.677090	0.659003	0.004326
dL49	1.245966	0.604738	0.006564
dL66	0.829367	0.327805	0.013901
dL68	1.905262	0.985186	0.019118
dL85	2.252862	1.273455	0.004588
dL88	1.738547	0.993421	0.022737
dL93	0.878960	0.799635	0.027271
dL96	2.046738	1.131799	0.006990
dL99	1.379210	0.718728	0.006467
dL104	1.206292	0.486265	0.015762
dL113	1.942965	0.842184	0.032765
dL118	1.125657	1.238468	0.026087
dL122	1.475625	0.718894	0.015184
dL126	1.520173	0.606880	0.012469
dL130	0.722666	0.576500	0.042083
dL134	2.160132	0.910095	0.020057
dL140	0.856860	0.693956	0.023048
dL148	1.012987	0.360663	0.050073
dL153	1.931629	0.988875	0.018785
dL165	0.772153	0.172524	0.016511
dL181	1.207209	0.231786	0.009111
dL184	1.175426	0.740175	0.008933
dL194	0.797176	1.078234	0.011772
dL197	0.866544	-0.356430	0.010751
dL199	0.915120	0.436022	0.031613
dL200	1.526600	0.658183	0.004873

Note. Higher values of *a* (.9 and above) indicate more discriminating items. Values of *b* are standardized (M=0, SD=1), and indicate where the item offered its maximum amount of information in relation to the theta continuum. Estimates of *b* near zero indicate that the item was most informative around the cutoff score.

Table 3

Item Numbers	Discrimination (*a*) parameters	Location (*b*) parameter	Guessing (*c*) parameters
dL8	0.980921	0.636035	0.028555
dL12	0.959749	0.688409	0.036879
dL14	1.146741	0.448451	0.084332
dL24	1.136186	0.562942	0.028030
dL28	0.885771	-0.710296	0.057590
dL29	1.031435	0.992918	0.023174
dL33	0.956050	0.101145	0.061335
dL38	0.766614	-0.046754	0.028577
dL48	1.027193	0.556718	0.028181
dL50	1.209125	0.530358	0.142482
dL55	0.977172	0.390924	0.026953
dL56	0.898090	-0.094946	0.180389
dL64	0.843086	0.382024	0.035551
dL73	1.263052	0.028980	0.018730
dL76	0.953530	0.095594	0.093521
dL78	1.293527	-0.018921	0.026868
dL80	1.950157	0.581470	0.027093
dL87	1.942796	0.695162	0.015134
dL89	1.271332	0.669840	0.087425
dL92	1.210384	0.598580	0.074629
dL95	1.709613	0.598749	0.023021
dL98	1.198796	0.265935	0.188871
dL103	1.015673	0.754834	0.034204
dL107	1.442159	0.184698	0.081480
dL110	1.465250	0.509566	0.100851
dL114	1.311050	0.860459	0.095367
dL120	0.788095	0.342749	0.026322
dL121	1.029134	0.267312	0.022794
dL125	1.859439	0.831850	0.020946
dL129	1.141084	0.340052	0.081760
dL137	1.004195	0.305975	0.043178
dL139	1.332545	0.802019	0.026167
dL143	0.809465	0.274840	0.040320
dL146	1.724311	0.632386	0.037915
dL160	0.764136	-0.161075	0.043215
dL166	1.802480	0.456331	0.047272
dL179	1.162085	0.062024	0.025194
dL190	1.155319	0.784416	0.054026

Note. Higher values of *a* (.9 and above) indicate more discriminating items. Values of *b* are standardized (M=0, SD=1), and indicate where the item offered its maximum amount of information in relation to the theta continuum. Estimates of *b* near zero indicate that the item was most informative around the cutoff score.

Table 4

Item Parameter Estimates for the New T-F Items

Item Numbers	Discrimination (*a*) parameters	Location (*b*) parameters	Guessing (*c*) parameters
dL17	0.908785	0.010691	0.049039
dL40	1.018393	-0.553529	0.076898
dL47	1.164695	0.052123	0.266363
dL62	0.859748	0.031719	0.031545
dL72	1.362154	-0.073375	0.034694
dL79	1.312926	-0.129546	0.043667
dL90	1.784032	-0.397593	0.029507
dL97	0.958254	-0.685473	0.076282
dL101	0.908963	-0.502694	0.074521
dL102	1.732952	-0.573316	0.037078
dL109	0.962642	0.226718	0.080108
dL111	1.412728	0.238839	0.066638
dL115	1.198839	-0.796043	0.124468
dL117	1.206721	-0.281646	0.084670
dL119	0.865427	-0.125814	0.072327
dL124	0.797143	-0.570491	0.046705
dL127	1.263401	-0.199946	0.025586
dL135	1.008809	-1.220628	0.068157
dL138	1.279191	-0.331848	0.037278
dL141	2.028799	-0.368549	0.054888
dL144	1.393526	-0.396675	0.042690
dL147	1.161611	-0.043004	0.059215
dL149	0.971775	-0.350375	0.046221
dL150	0.888413	-0.356350	0.065749
dL152	0.933402	-0.487372	0.083079
dL155	1.846032	-0.348759	0.030276
dL157	1.178934	-0.297098	0.039538
dL162	1.168213	-0.218992	0.048662
dL164	1.572488	-0.354682	0.035465
dL167	1.545905	-0.396644	0.054967
dL169	1.083559	-0.266020	0.063106
dL172	1.087227	-0.690181	0.062445
dL178	1.567241	-0.302151	0.033803
dL193	1.047752	0.399057	0.061176
dL198	0.800454	0.181197	0.039819

Note. Higher values of *a* (.9 and above) indicate more discriminating items. Values of *b* are standardized (M=0, SD=1), and indicate where the item offered its maximum amount of information in relation to the theta continuum. Estimates of *b* near zero indicate that the item was most informative around the cutoff score.

Table 5

Item Parameter Estimates for the New J-P Items

Item Numbers	Discrimination (*a*) parameters	Location (*b*) parameters	Guessing (*c*) parameters
dL9	0.782191	-0.297457	0.044702
dL20	1.704985	0.471972	0.022973
dL23	0.737800	-0.054308	0.116174
dL26	1.876838	0.513410	0.017116
dL30	0.836613	0.833042	0.016542
dL32	1.303193	-0.120130	0.017698
dL41	1.487536	-0.229766	0.044019
dL43	0.680165	-0.033073	0.034555
dL53	1.446886	0.239033	0.028139
dL69	0.952367	-0.752042	0.042549
dL74	2.315338	0.551211	0.017952
dL75	0.936237	-0.052866	0.023683
dL83	0.860384	0.859384	0.008389
dL86	1.647835	-0.055462	0.026016
dL91	0.773768	0.521192	0.062579
dL94	0.727961	-0.222998	0.033388
dL100	0.743922	-0.078738	0.026170
dL116	1.054706	-0.479587	0.021304
dL123	0.820528	0.846044	0.031247
dL132	1.813928	0.195195	0.032865
dL142	1.786474	-0.148566	0.033523
dL156	1.406017	0.657907	0.012351
dL163	0.778171	-0.770853	0.050804
dL176	2.219276	0.482311	0.004851
dL180	1.784546	0.601132	0.029353
dL183	0.833711	0.172231	0.041872
dL185	1.659617	0.332204	0.019782
dL187	0.878284	0.700646	0.032055
dL189	1.443279	0.118255	0.035079
dL192	1.260243	-0.185034	0.022813

Note. Higher values of *a* (.9 and above) indicate more discriminating items. Values of *b* are standardized (M=0, SD=1), and indicate where the item offered its maximum amount of information in relation to the theta continuum. Estimates of *b* near zero indicate that the item was most informative around the cutoff score.

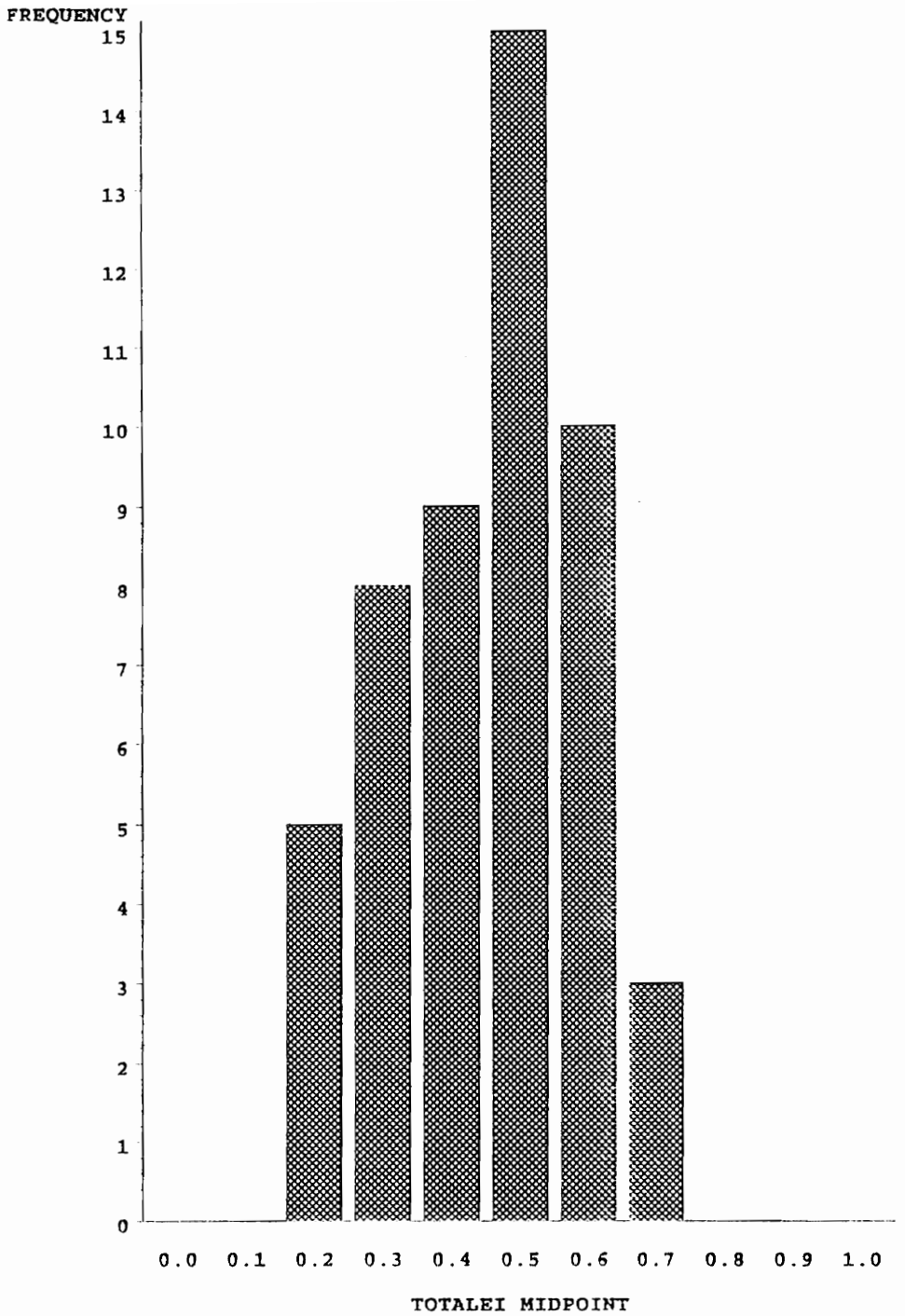

Figure 1: Item-Total Frequency Distribution for E-I Scale. Item-total frequency distribution between the new items and a 10-item composite consisting of the best items from the original MBTI for the E-I scale.

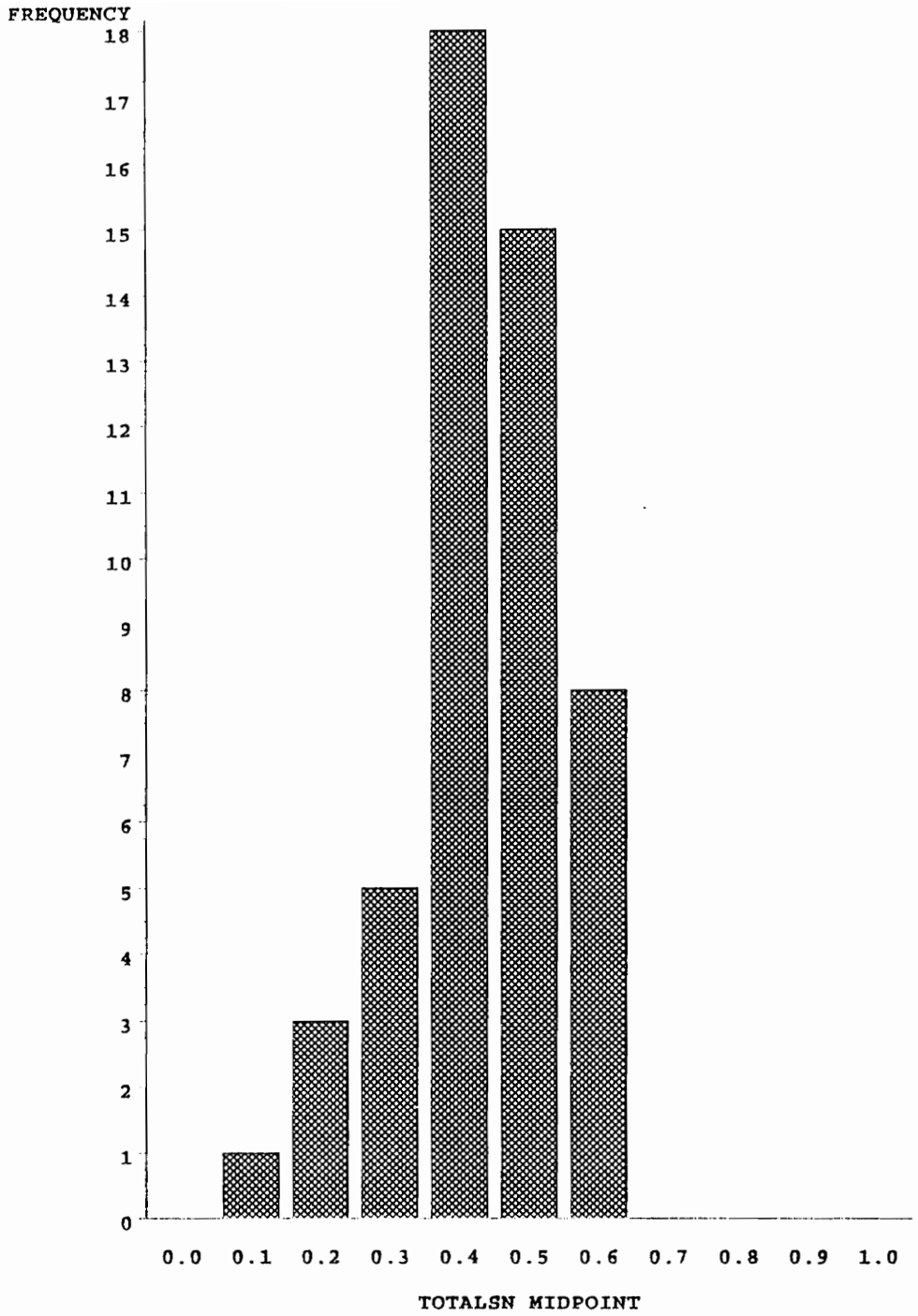

Figure 2: Item-Total Frequency Distribution for S-N Scale. Item-total frequency distribution between the new items and a 10-item composite consisting of the best items from original MBTI for the S-N scale.

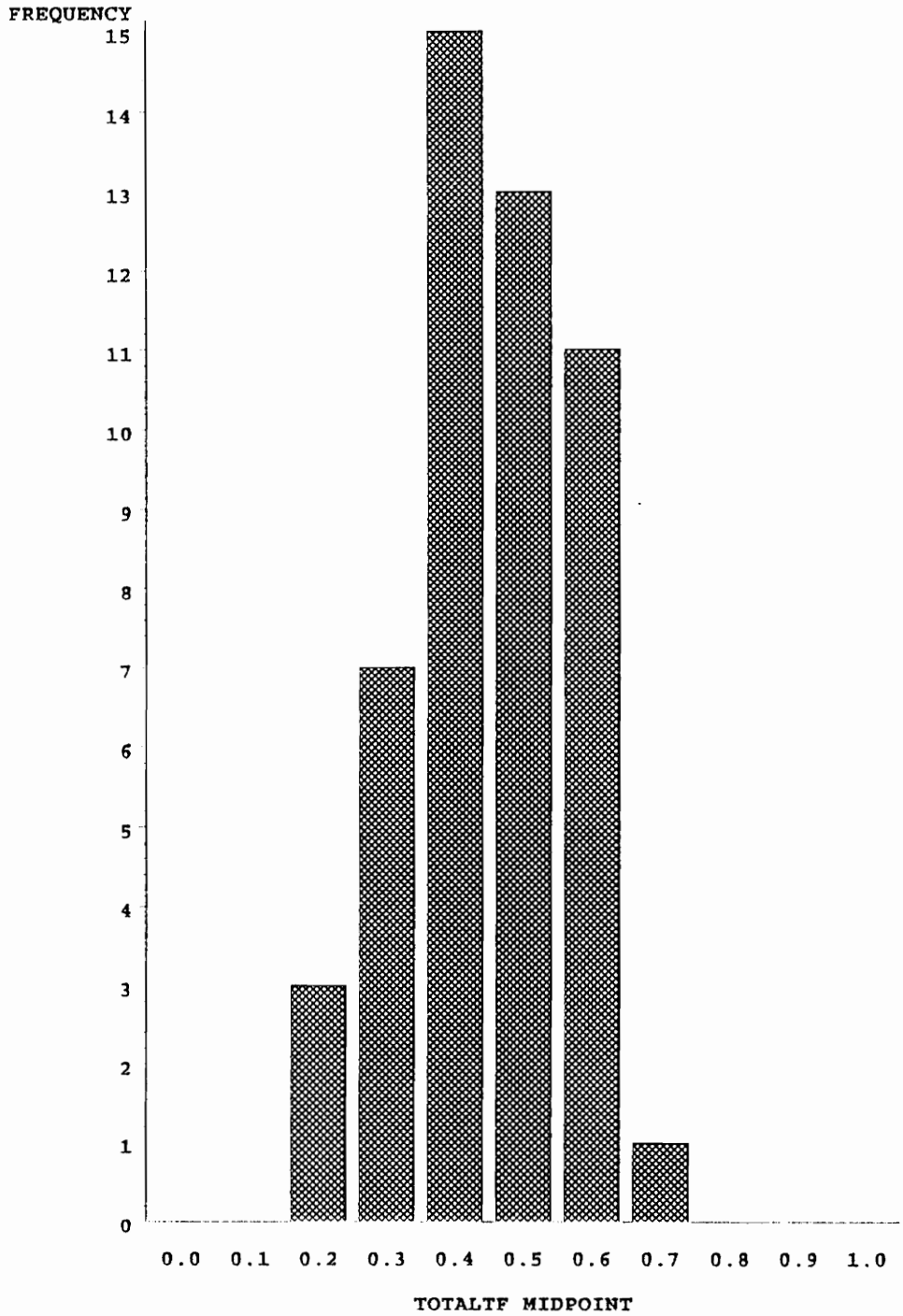

Figure 3: Item-Total Frequency Distribution for T-F Scale. Item-total frequency distribution between the new items and a 10-item composite consisting of the best items from original MBTI for the T-F scale.

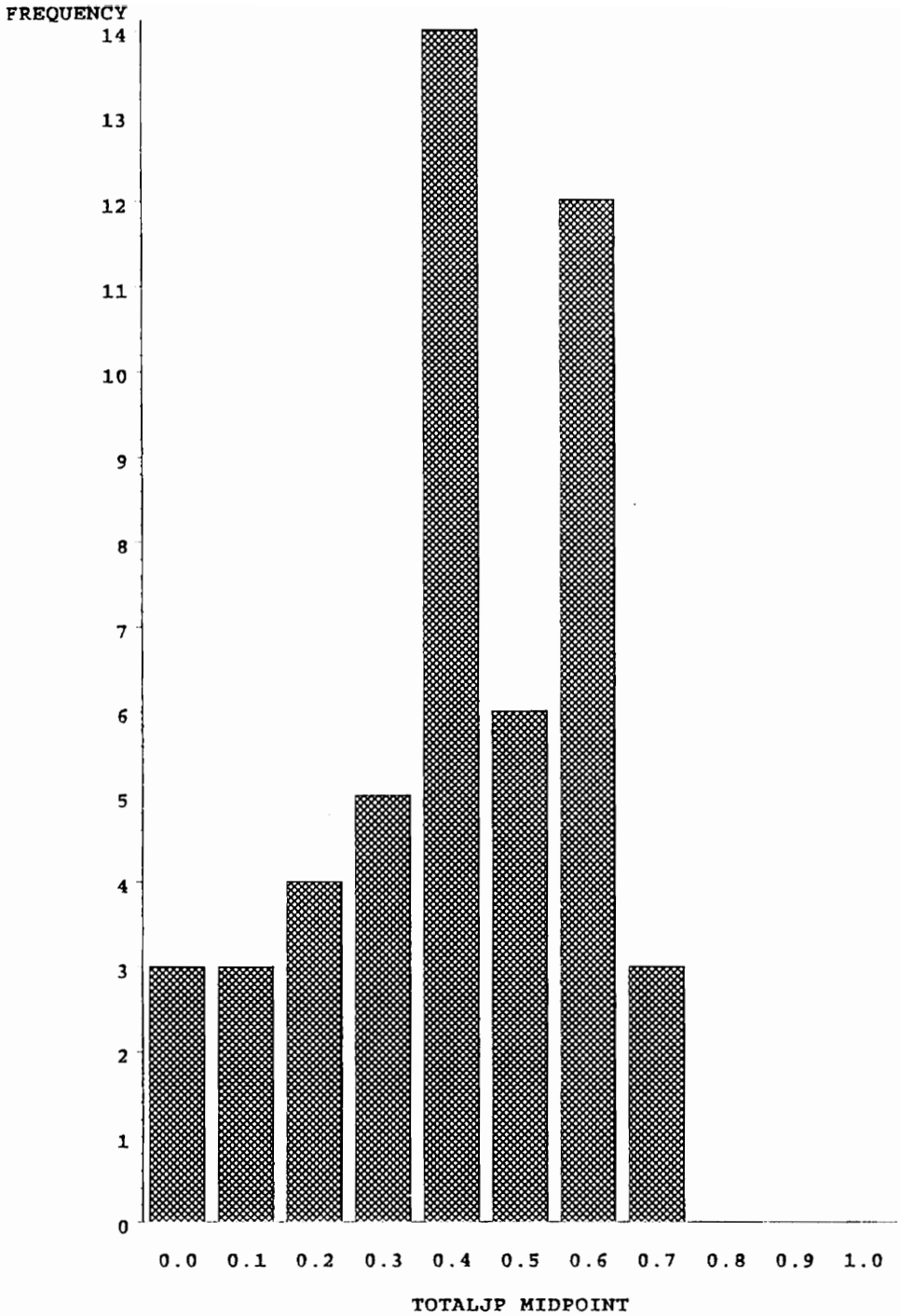

Figure 4: Item-Total Frequency Distribution for J-P Scale. Item-total frequency distribution between the new items and a 10-item composite consisting of the best items from original MBTI for the J-P scale.

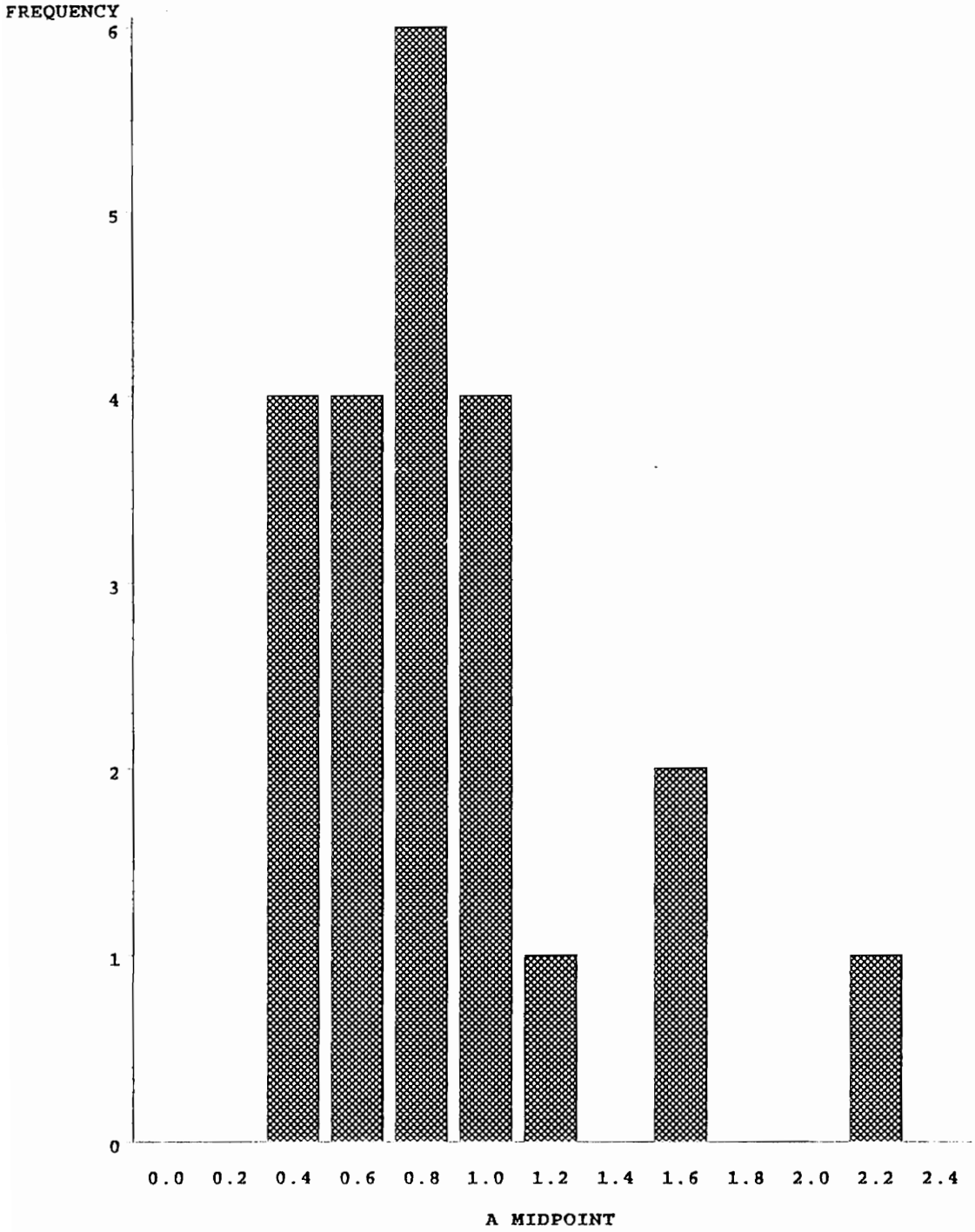

Figure 5: Frequency Distribution of the As: E-I Original Items. Frequency distributions of the discrimination (a) parameters for the original MBTI items for the E-I scale.

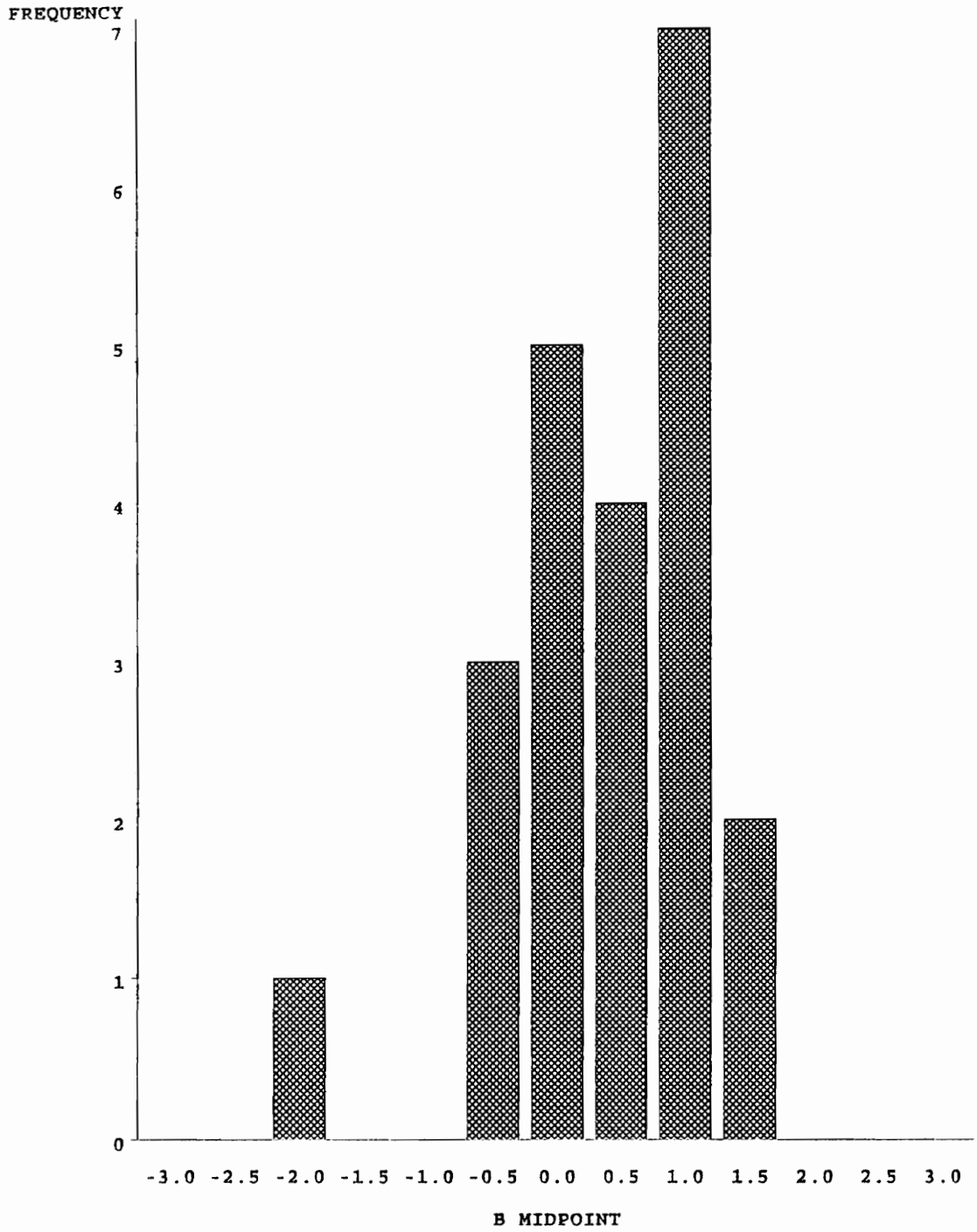

Figure 6: Frequency Distribution of the Bs: E-I Original Items. Frequency distributions of the difficulty (b) parameters for the original MBTI items for the E-I scale.

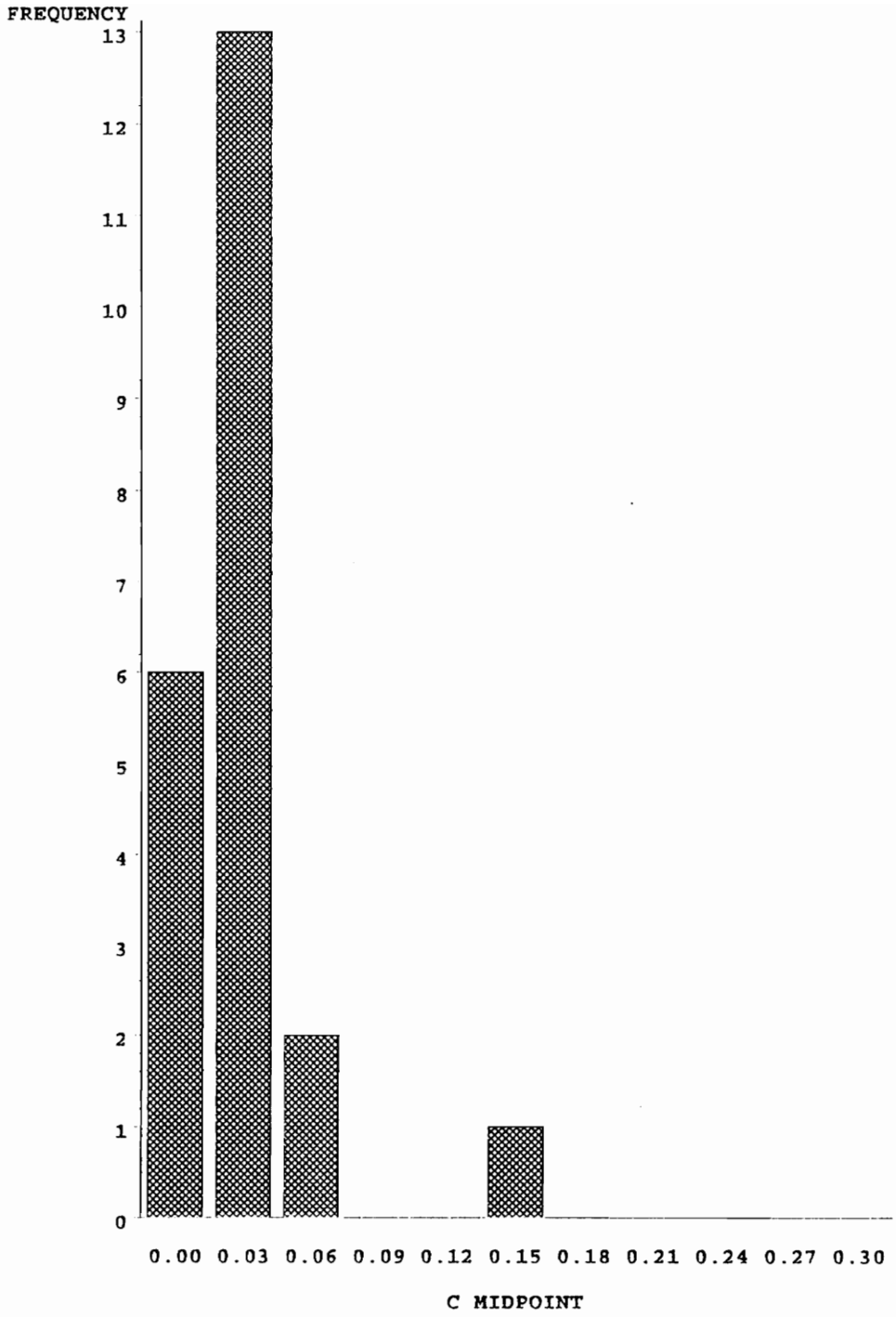

Figure 7: Frequency Distribution of the Cs: E-I Original Items. Frequency distributions of the pseudoguessing (c) parameters for the original MBTI items for the E-I scale.

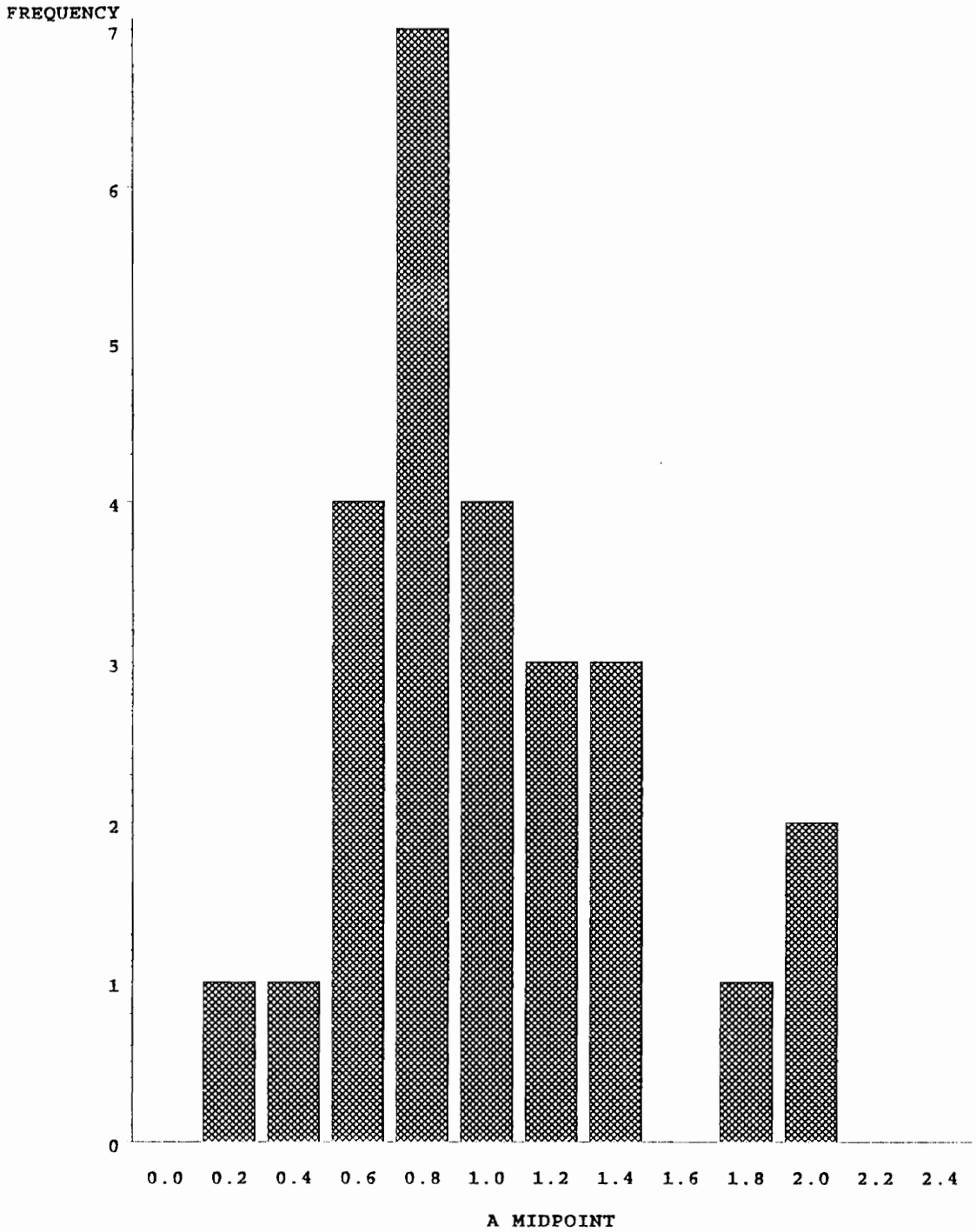

Figure 8: Frequency Distribution of the As: S-N Original Items.68 Frequency distributions of the discrimination (a) parameters for the original MBTI items for the S-N scale.

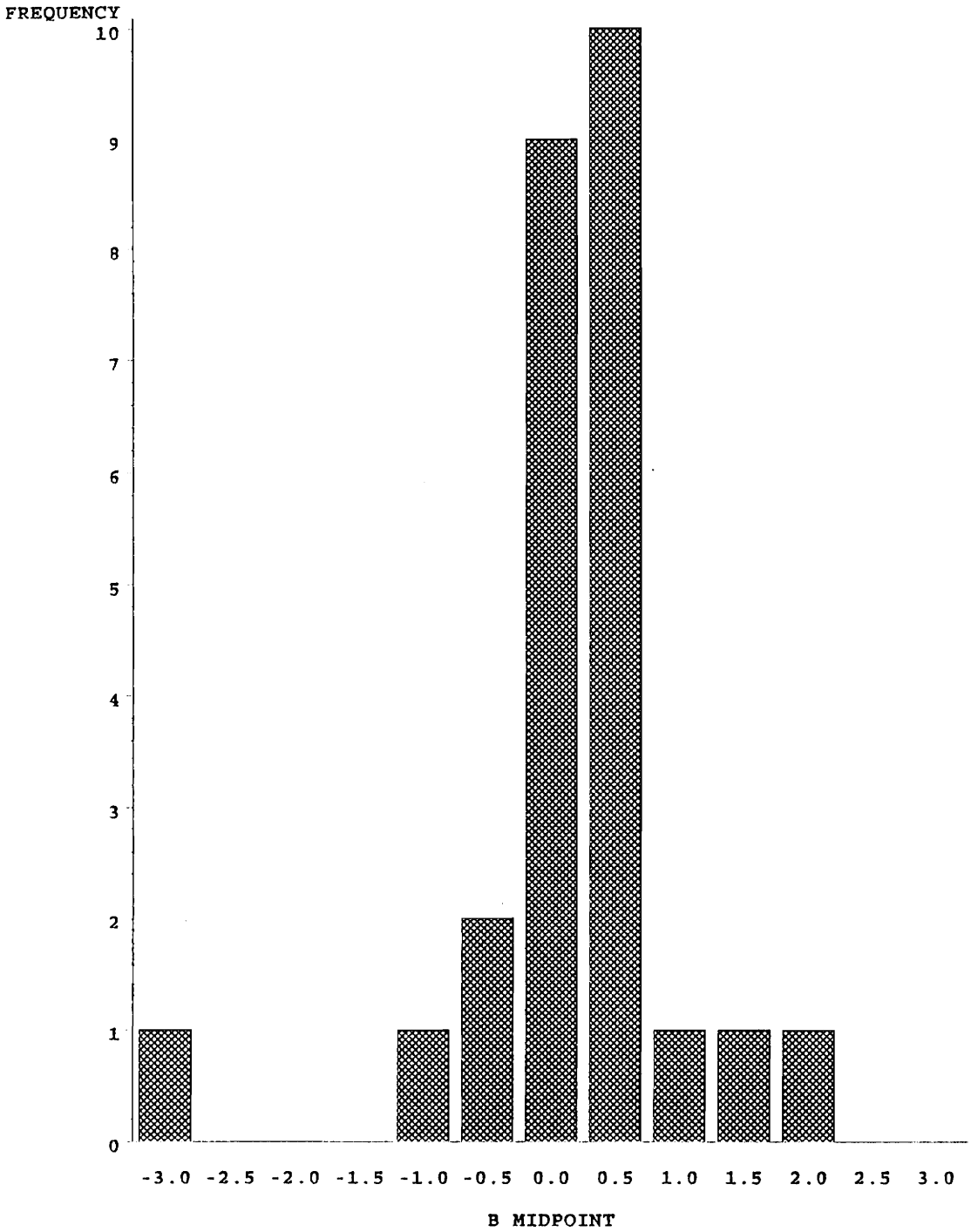

Figure 9: Frequency Distribution of the Bs: S-N Original Items. Frequency distributions of the difficulty (b) parameters for the original MBTI items for the S-N scale.

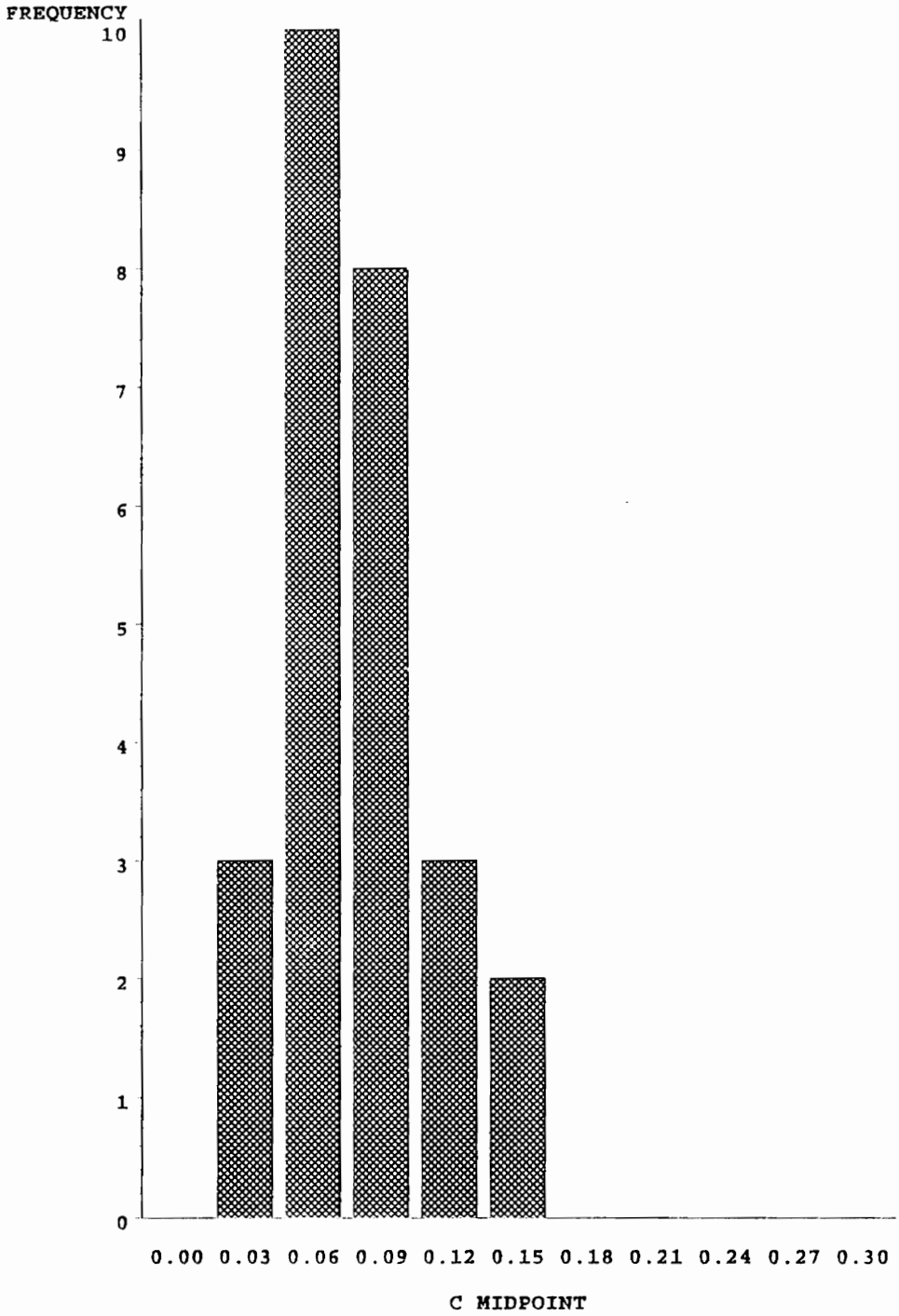

Figure 10: Frequency Distribution of the Cs: S-N Original Items. Frequency distributions of the pseudoguessing (c) parameters for the original MBTI items for the S-N scale.

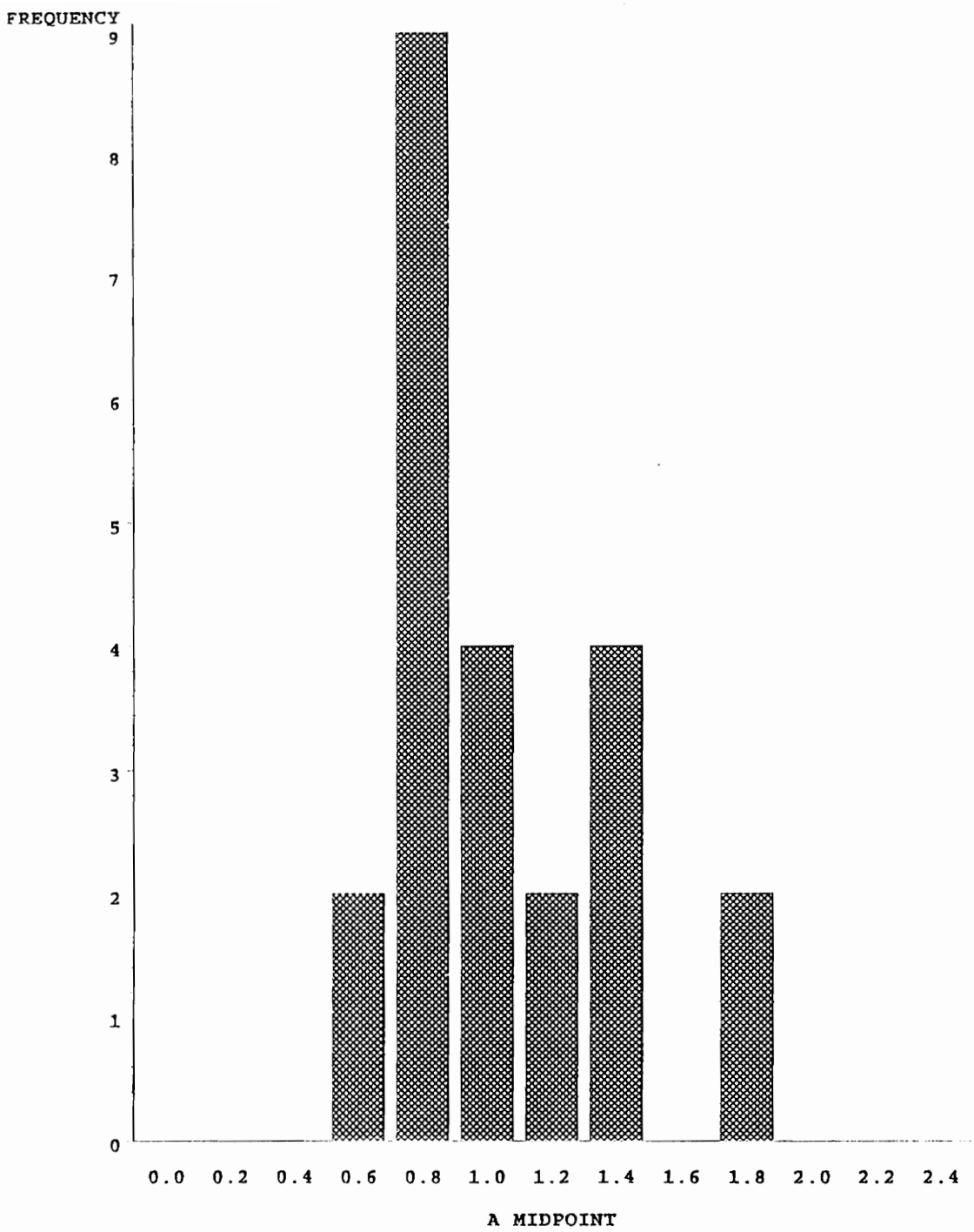

Figure 11: Frequency Distribution of the As: T-F Original Items. Frequency distributions of the discrimination (a) parameters for the original MBTI items for the T-F scale.

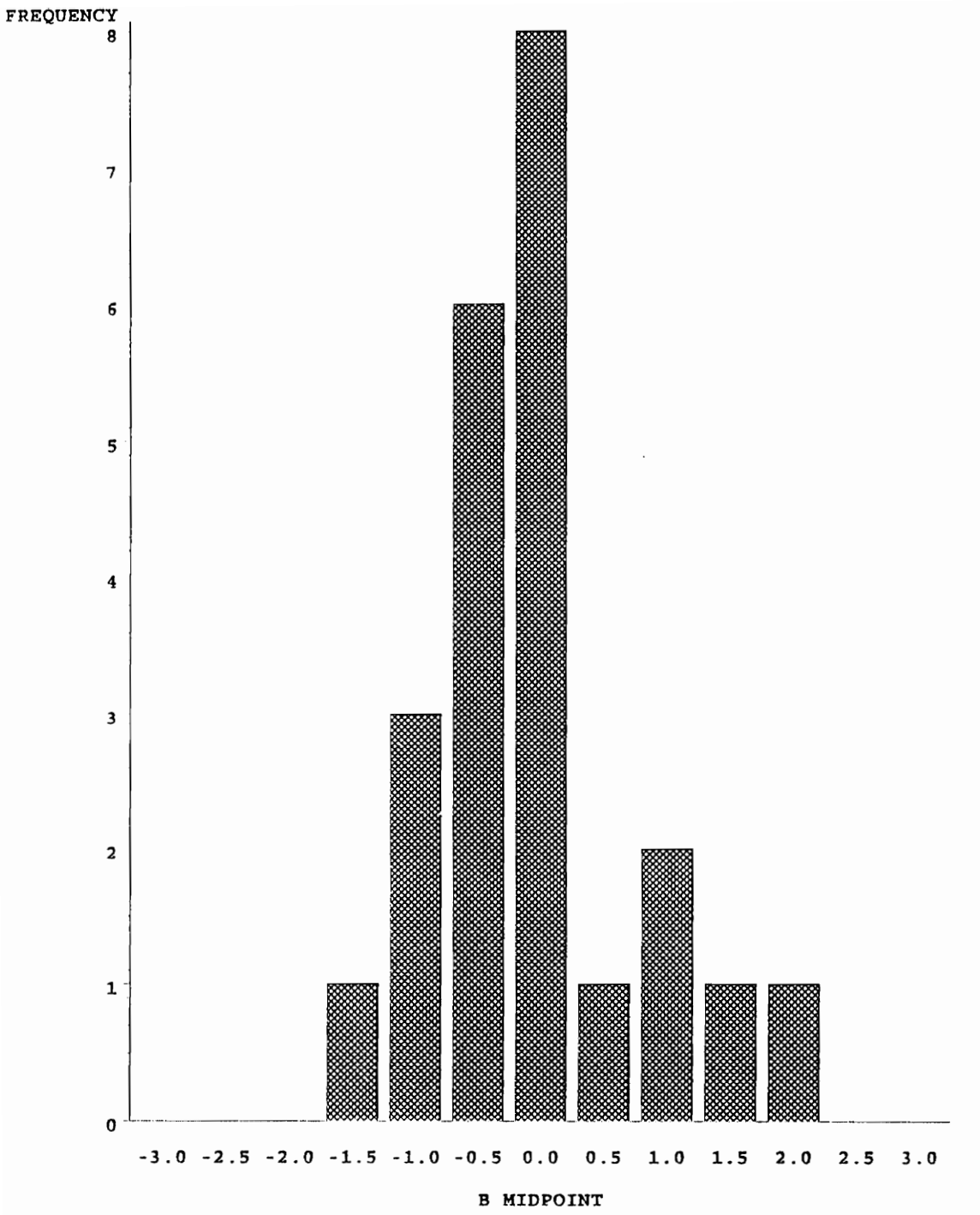

Figure 12: Frequency Distribution of the Bs: T-F Original Items. Frequency distributions of the difficulty(b) parameters for the original MBTI items for the T-F scale.

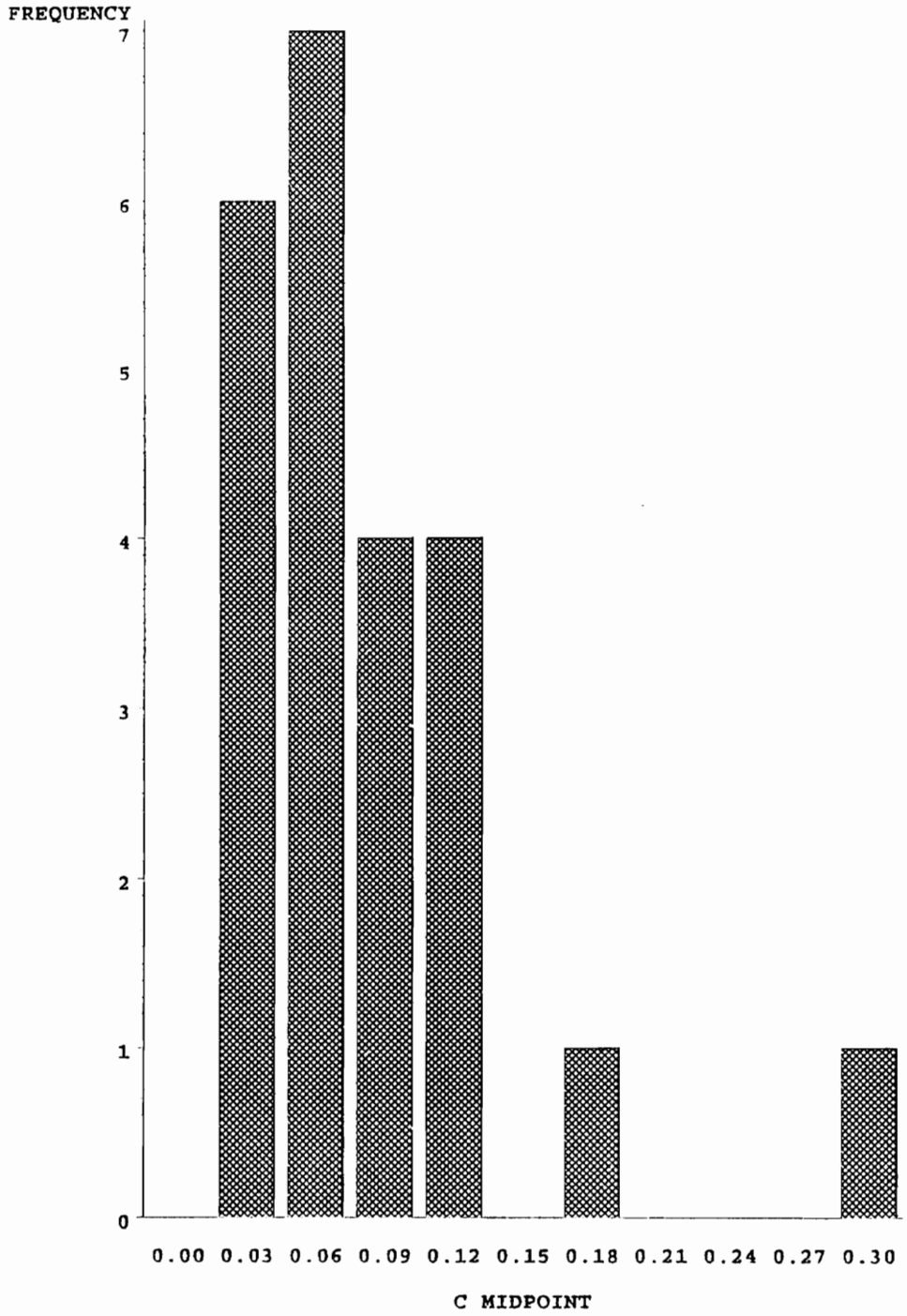

Figure 13: Frequency Distribution of the Cs: T-F Original Items. Frequency distributions of the pseudoguessing (c) parameters for the original MBTI items for the T-F scale.

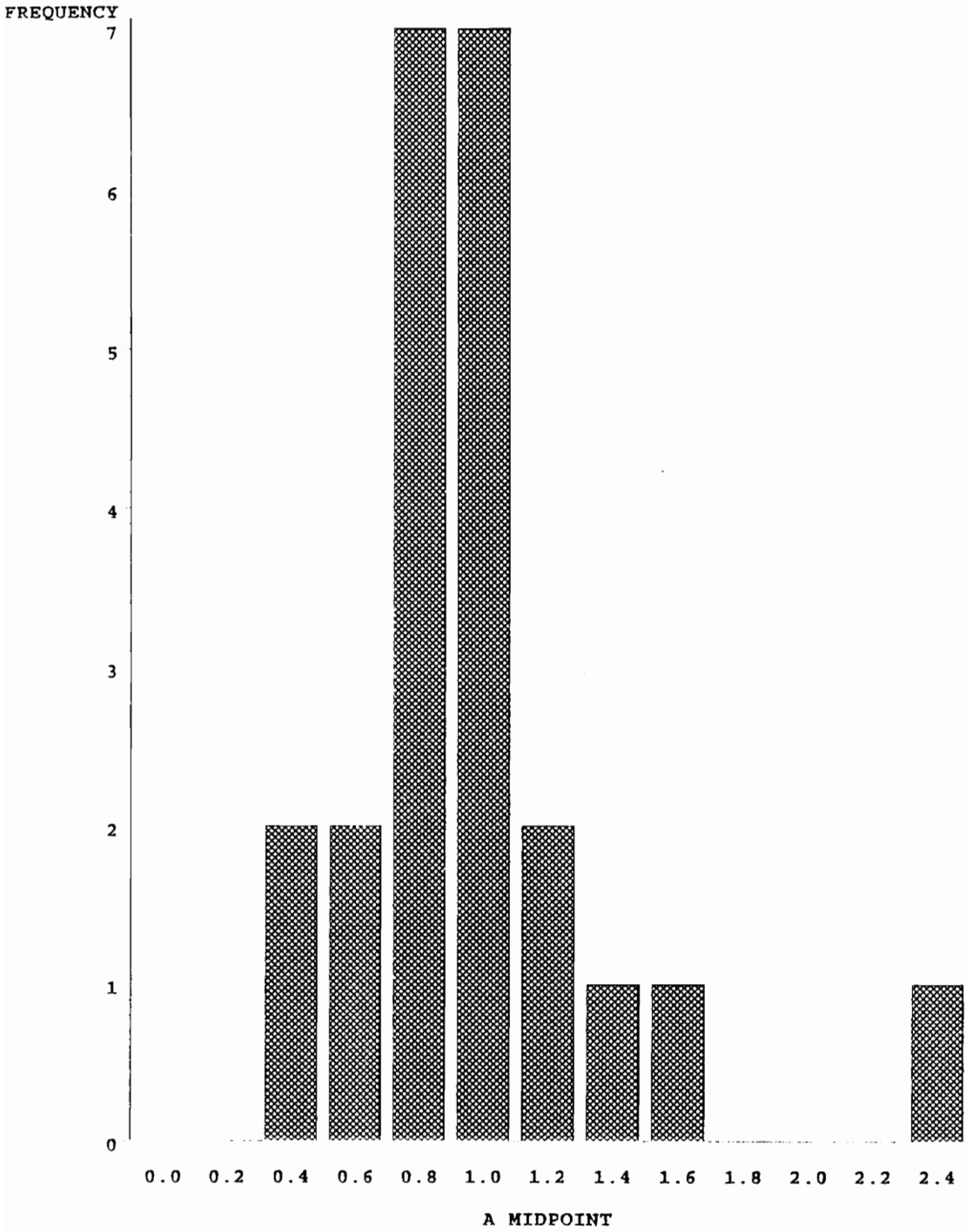

Figure 14: Frequency Distribution of the As: J-P Original Items. Frequency distributions of the discrimination (a) parameters for the original MBTI items for the J-P scale.

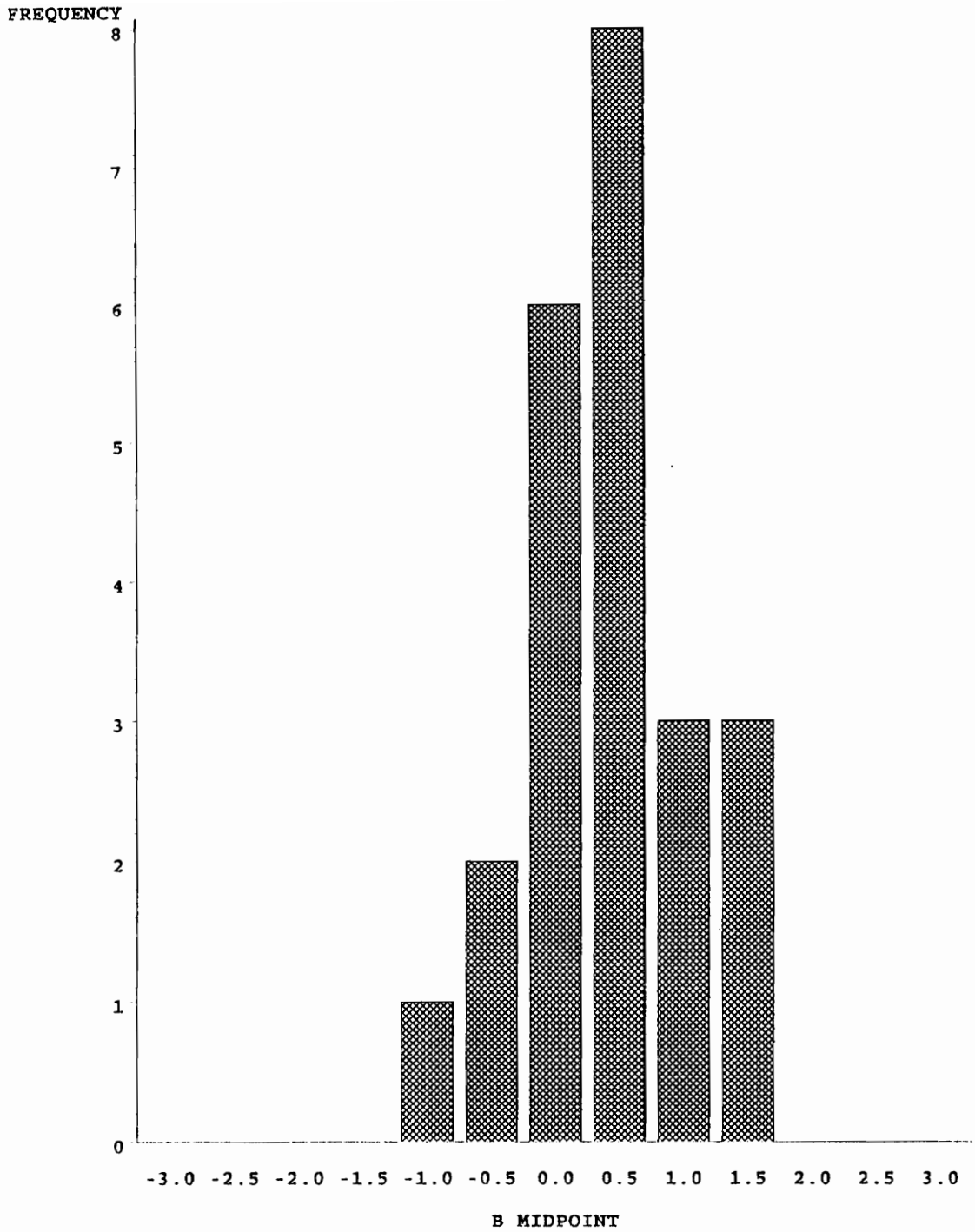

Figure 15: Frequency Distribution of the Bs: J-P Original Items. Frequency distributions of the difficulty (b) parameters for the original MBTI items for the J-P scale.

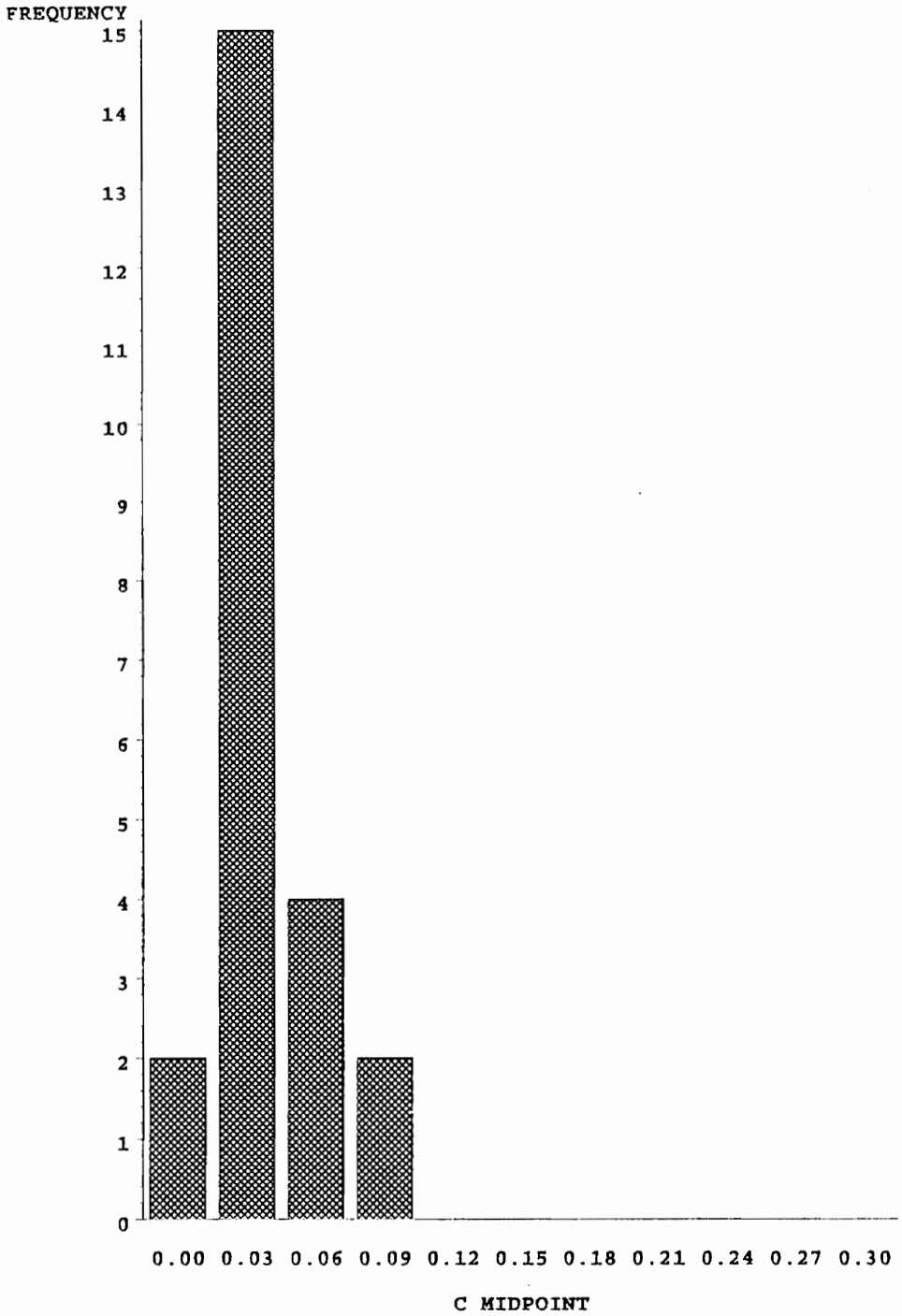

Figure 16: Frequency Distribution of the Cs: J-P Original Items. Frequency distributions of the pseudoguessing (c) parameters for the original MBTI items for the J-P scale.

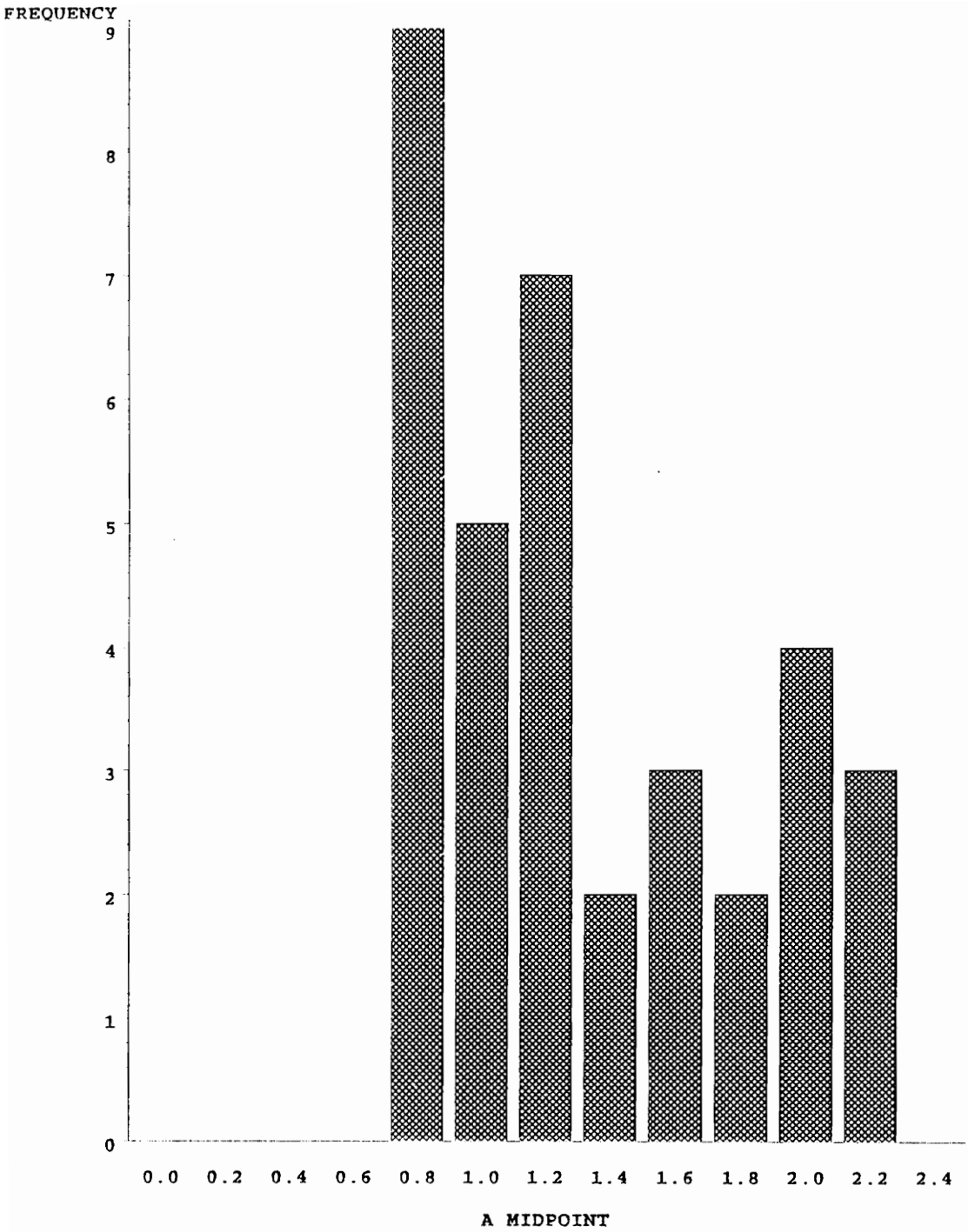

Figure 17: Frequency Distribution of the As: E-I New Items. Frequency distributions of the discrimination (a) parameters for all the new items for the E-I scale.

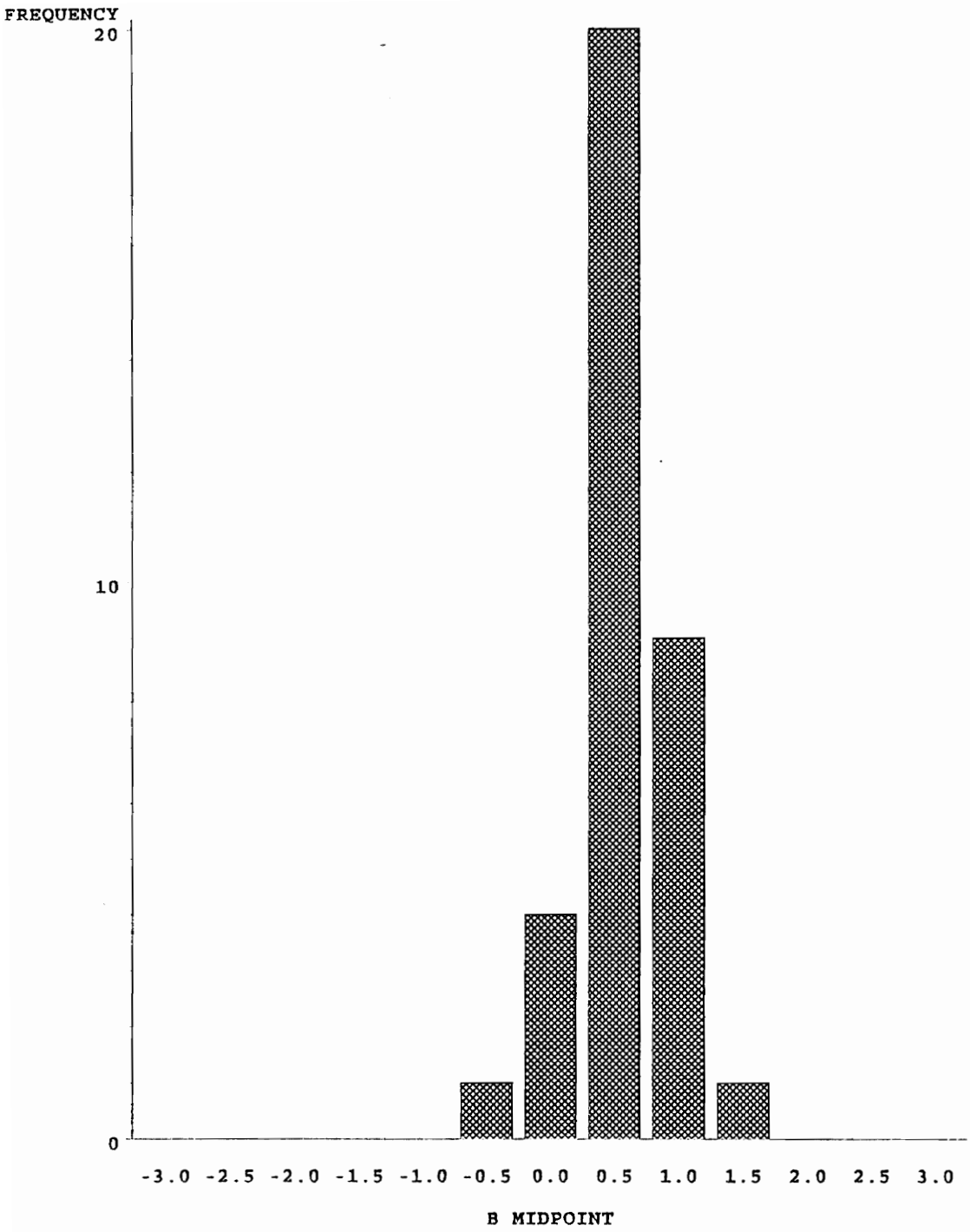

Figure 18: Frequency Distribution of the Bs: E-I New Items. Frequency distributions of the difficulty (b) parameters for all the new items for the E-I scale.

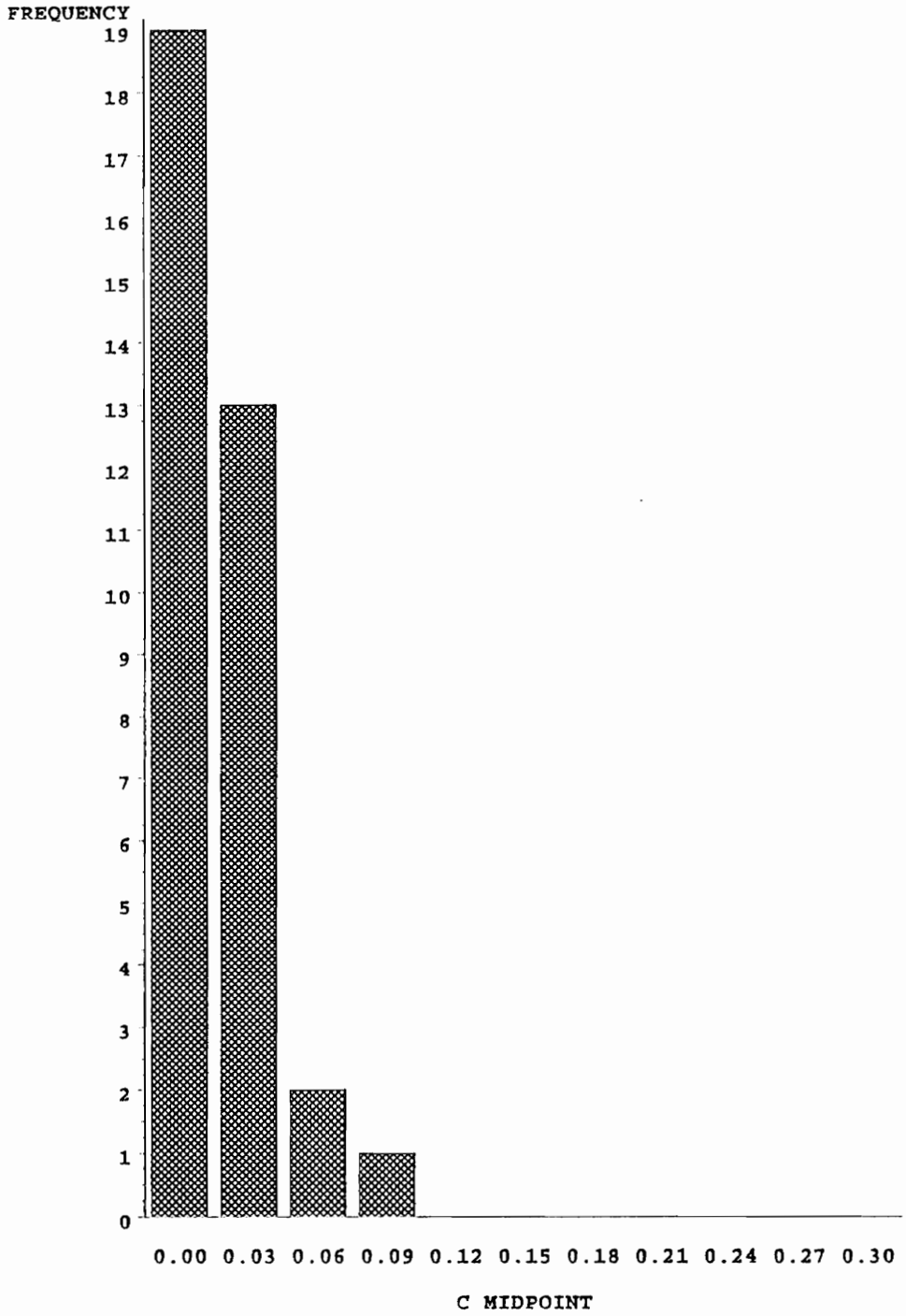

Figure 19: Frequency Distribution of the Cs: E-I New Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items for the E-I scale.

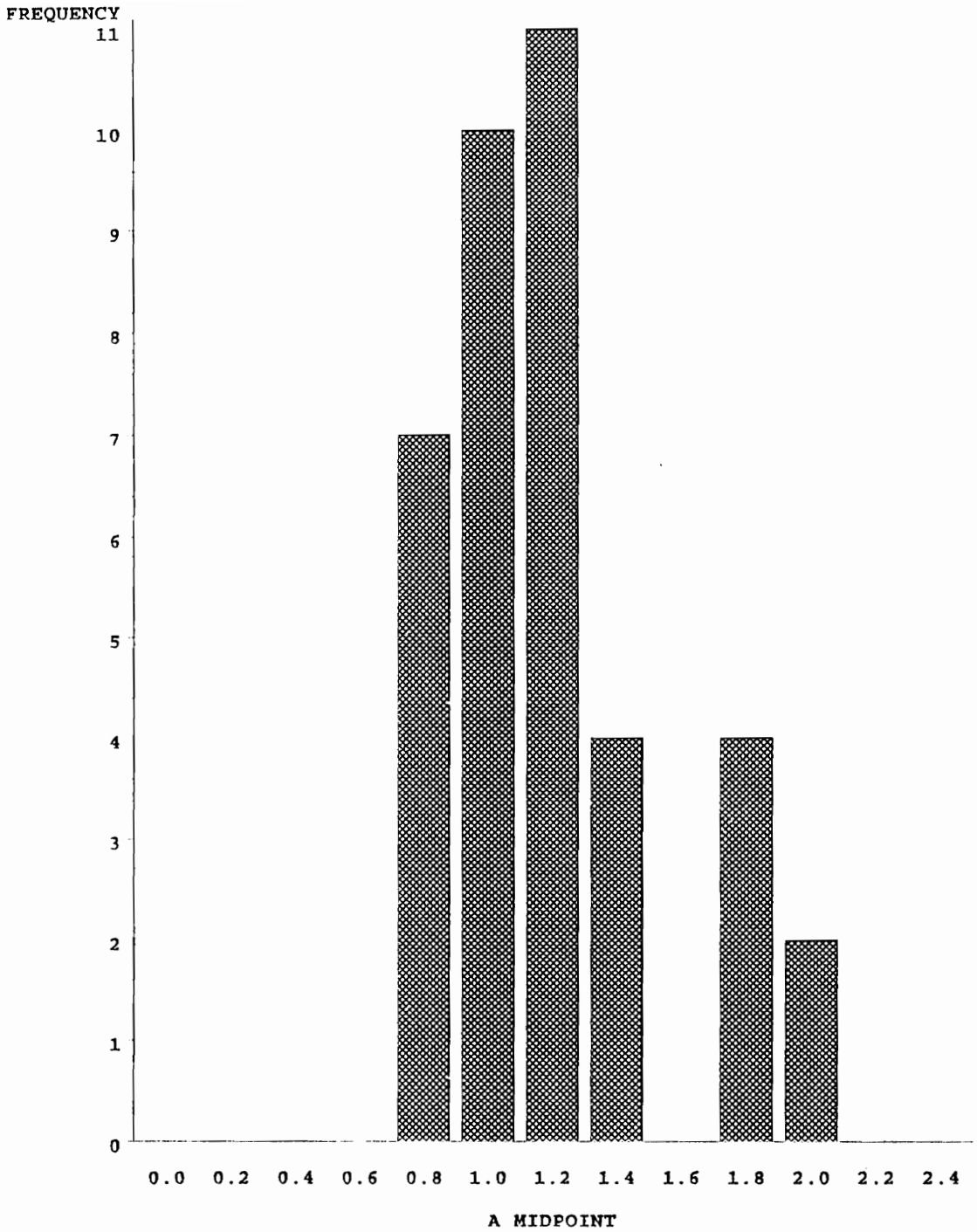

Figure 20: Frequency Distribution of the As: S-N New Items. Frequency distributions of the discrimination (a) parameters for all the new items for the S-N scale.

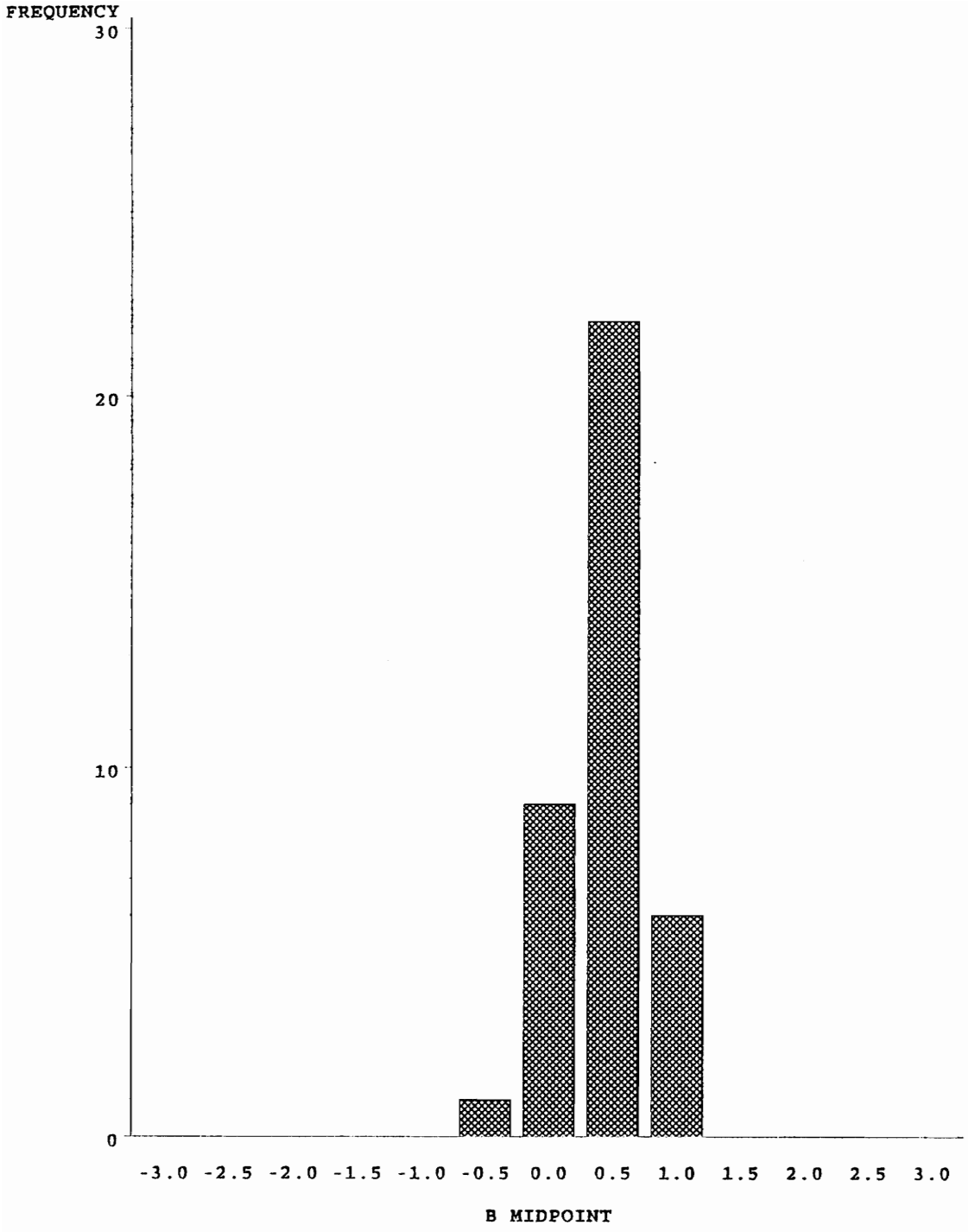

Figure 21: Frequency Distribution of the Bs: S-N New Items. Frequency distributions of the difficulty (b) parameters for all the new items for the S-N scale.

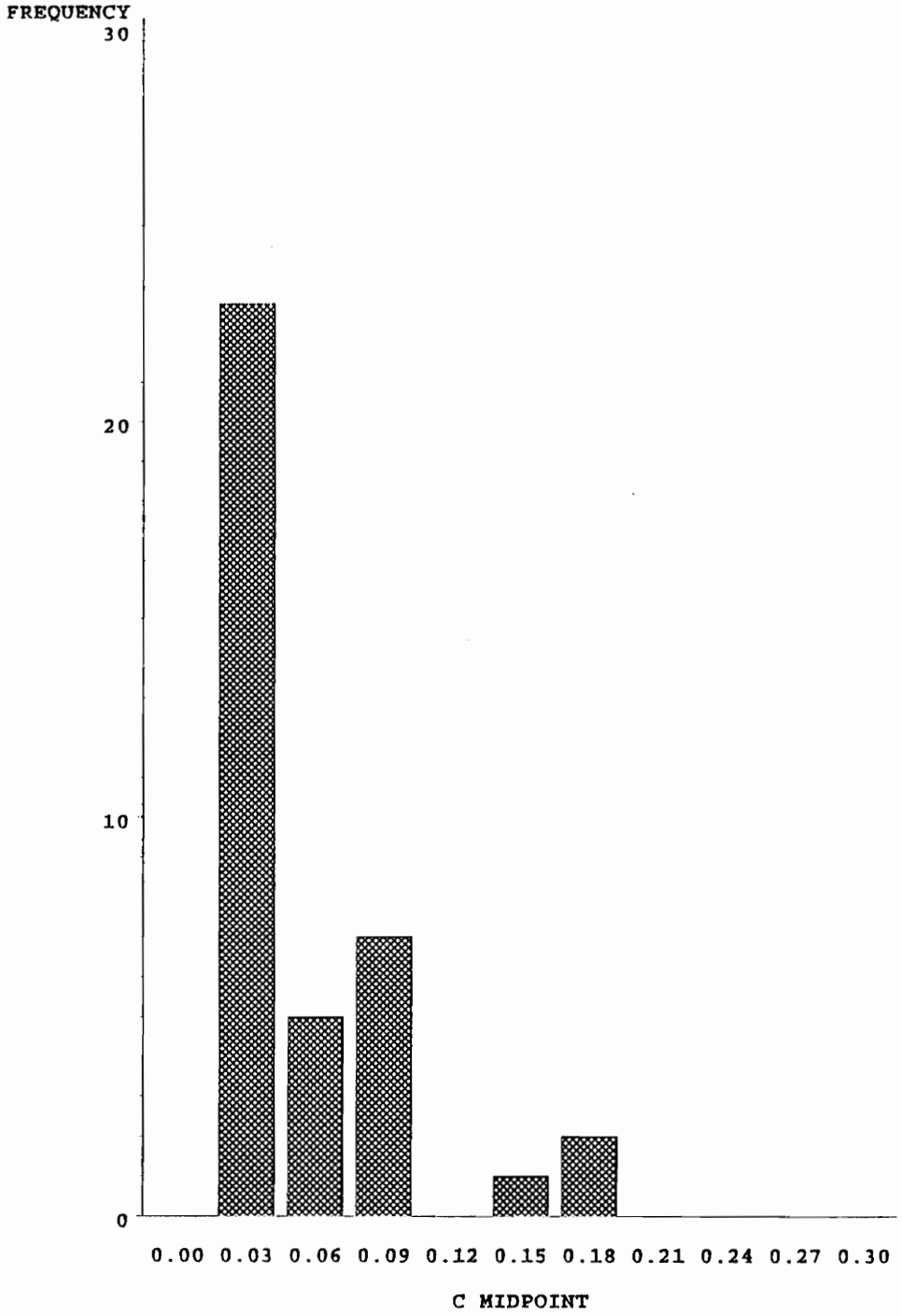

Figure 22: Frequency Distribution of the Cs: S-N New Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items for the S-N scale.

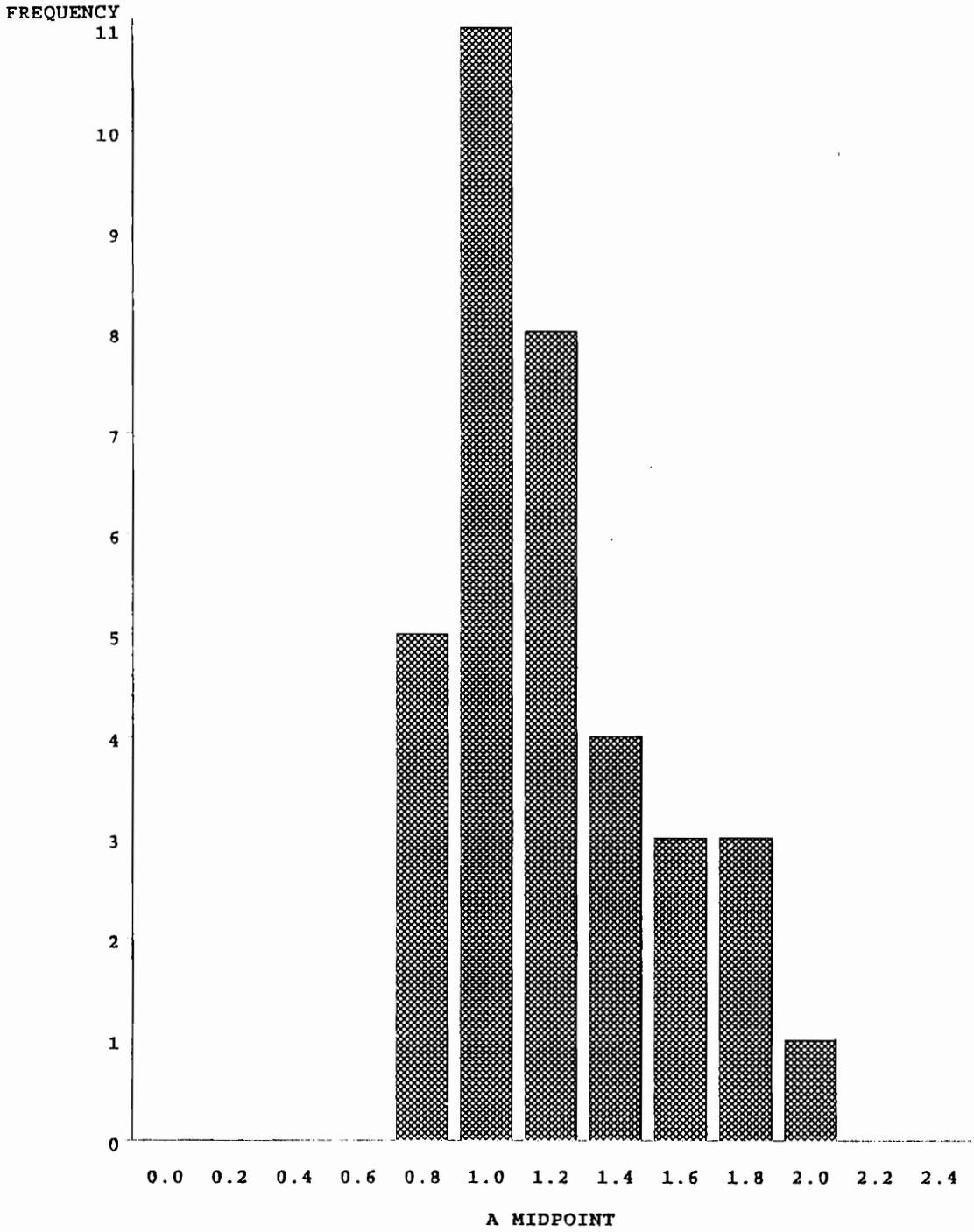

Figure 23: Frequency Distribution of the As: T-F New Items. Frequency distributions of the discrimination (a) parameters for all the new items for the T-F scale.

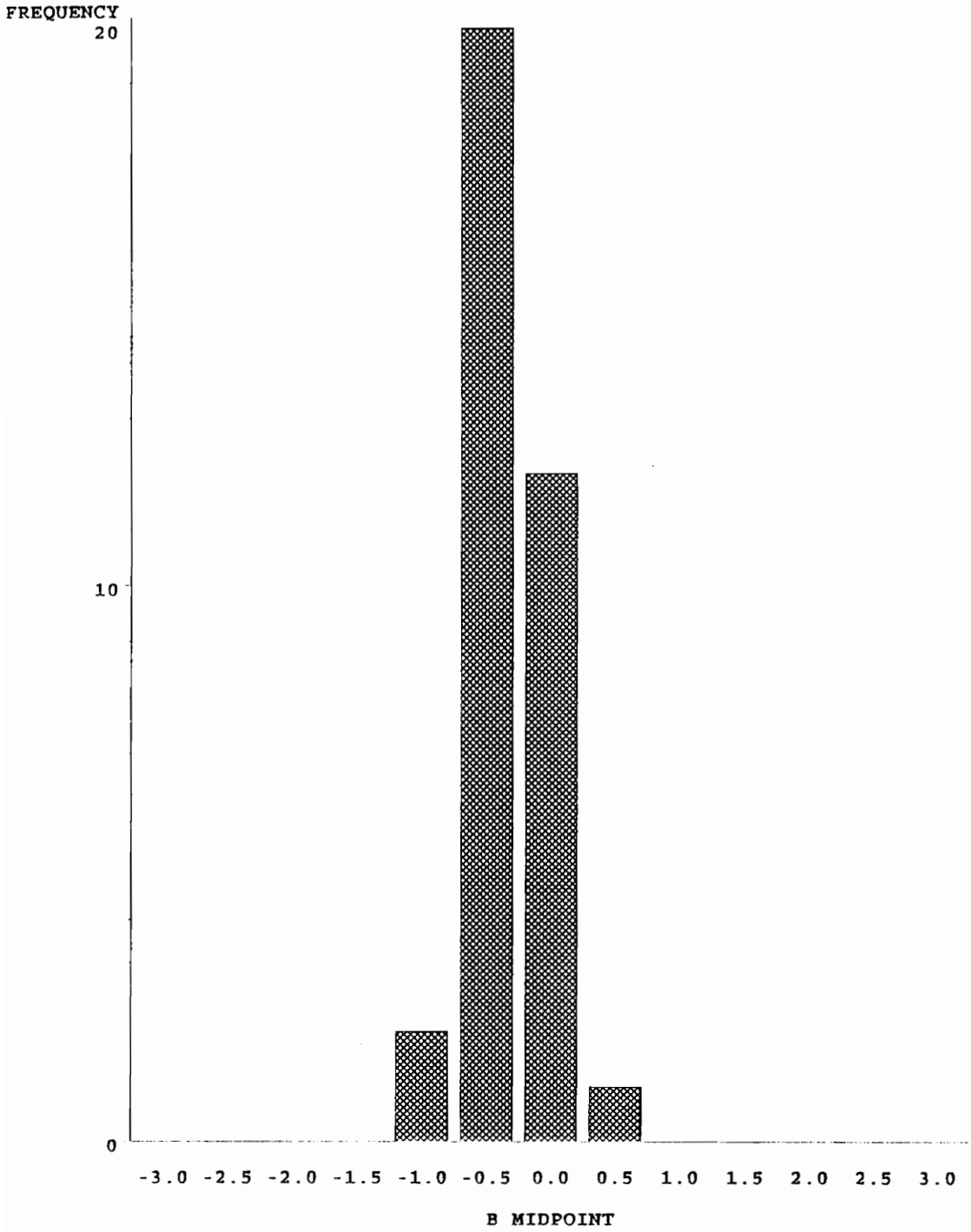

Figure 24: Frequency Distribution of the Bs: T-F New Items. Frequency distributions of the difficulty (b) parameters for all the new items for the T-F scale.

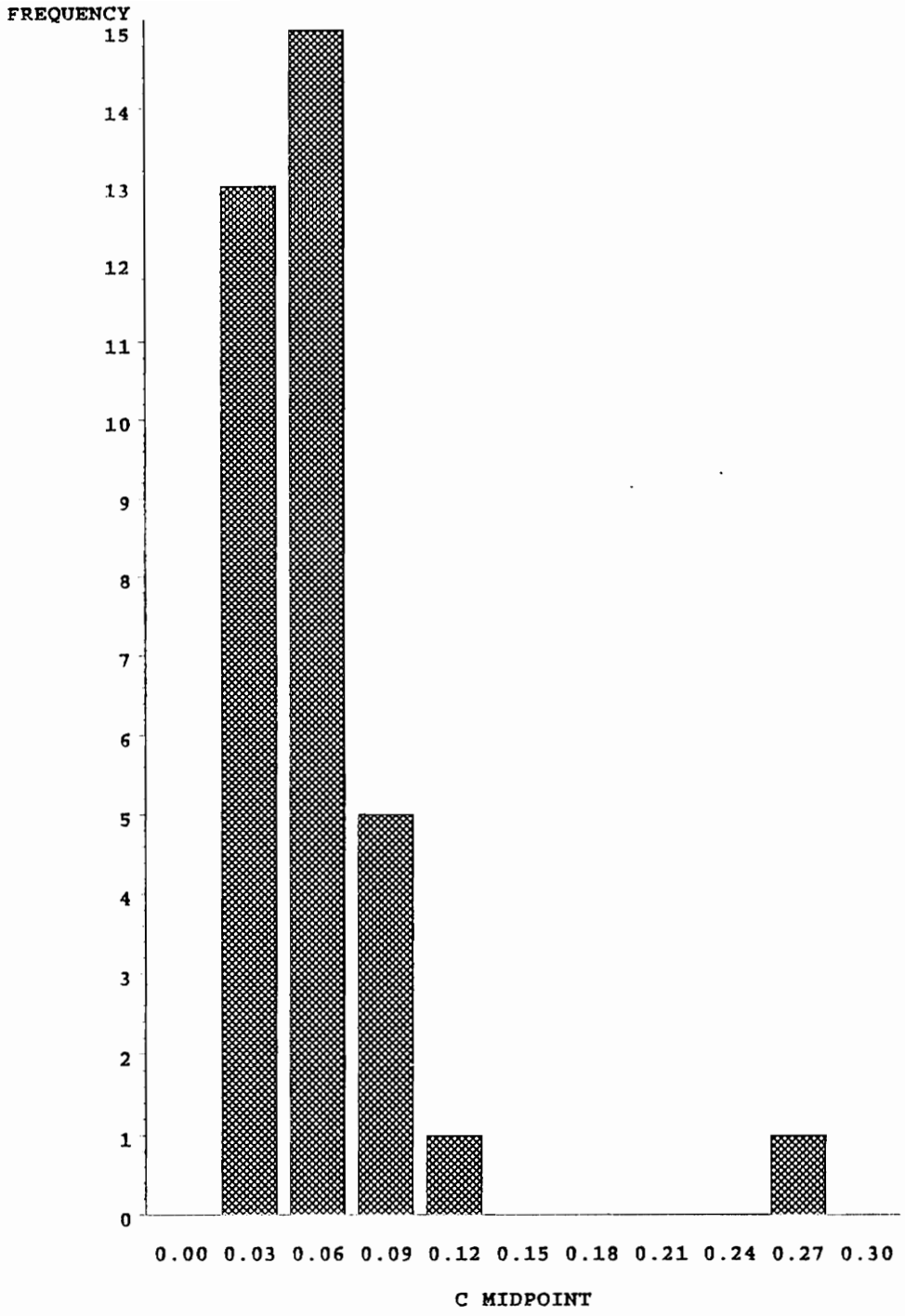

Figure 25: Frequency Distribution of the Cs: T-F New Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items for the T-F scale.

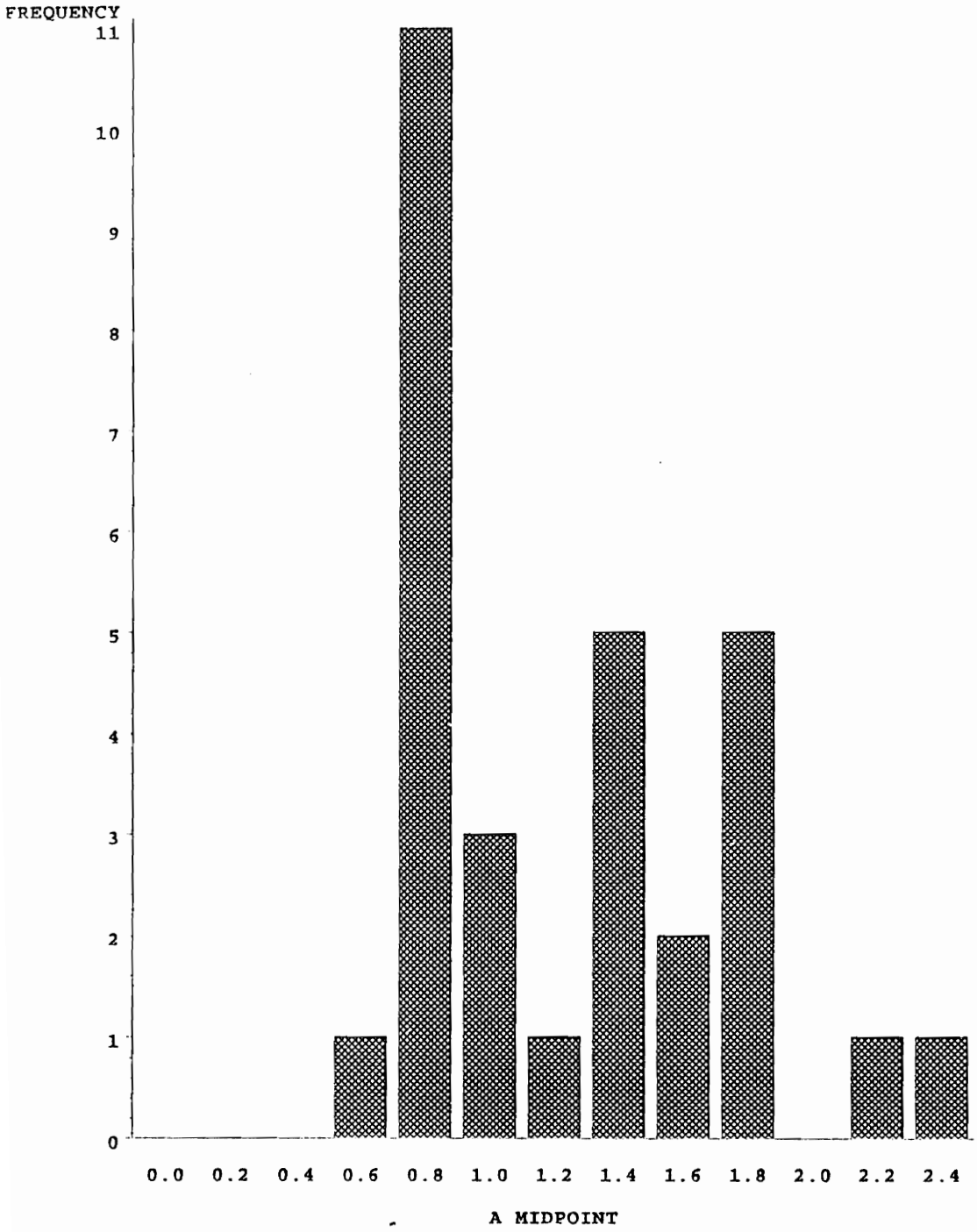

Figure 26: Frequency Distribution of the As: J-P New Items. Frequency distributions of the discrimination (a) parameters for all the new items for the J-P scale.

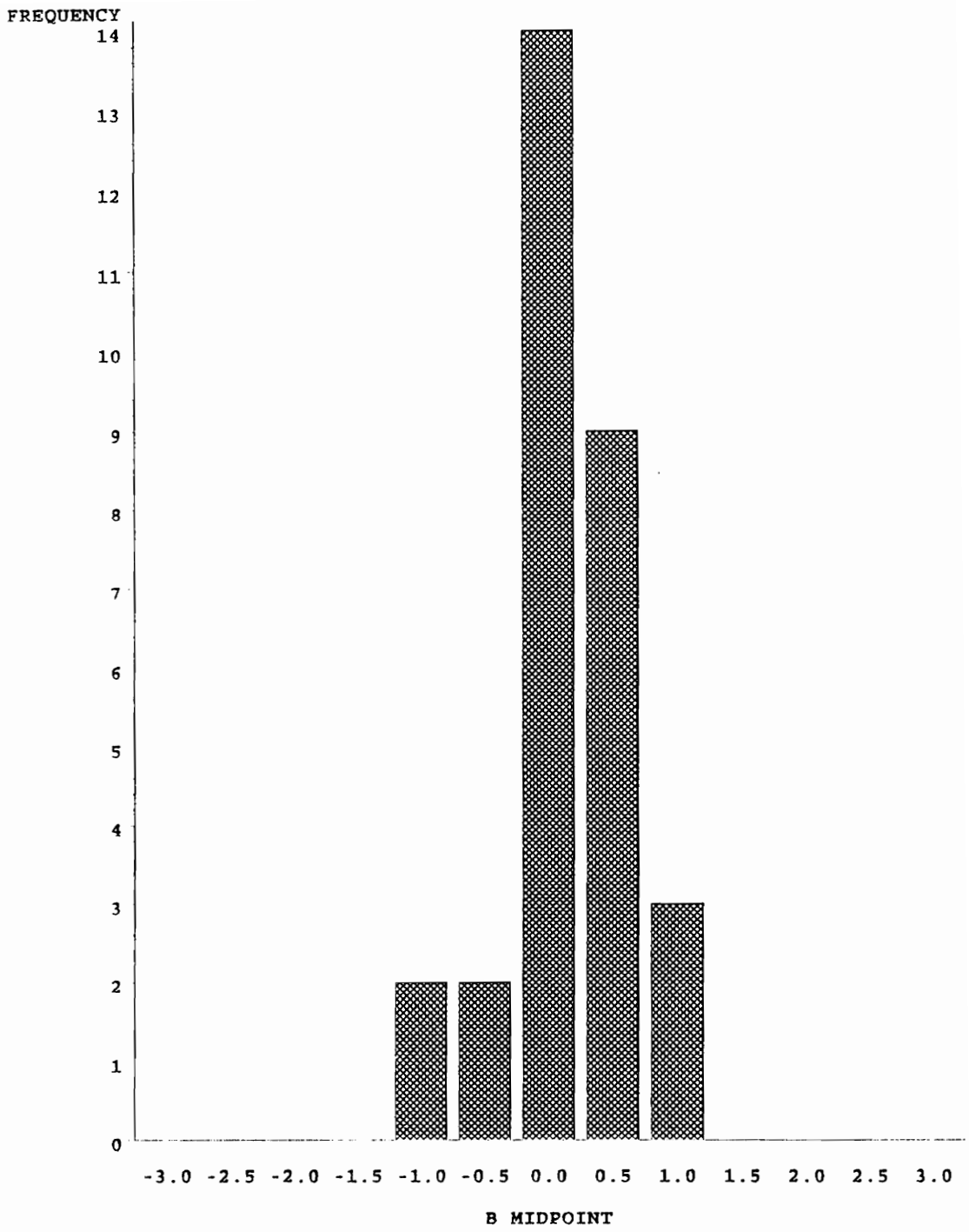

Figure 27: Frequency Distribution of the Bs: J-P New Items. Frequency distributions of the difficulty (b) parameters for all the new items for the J-P scale.

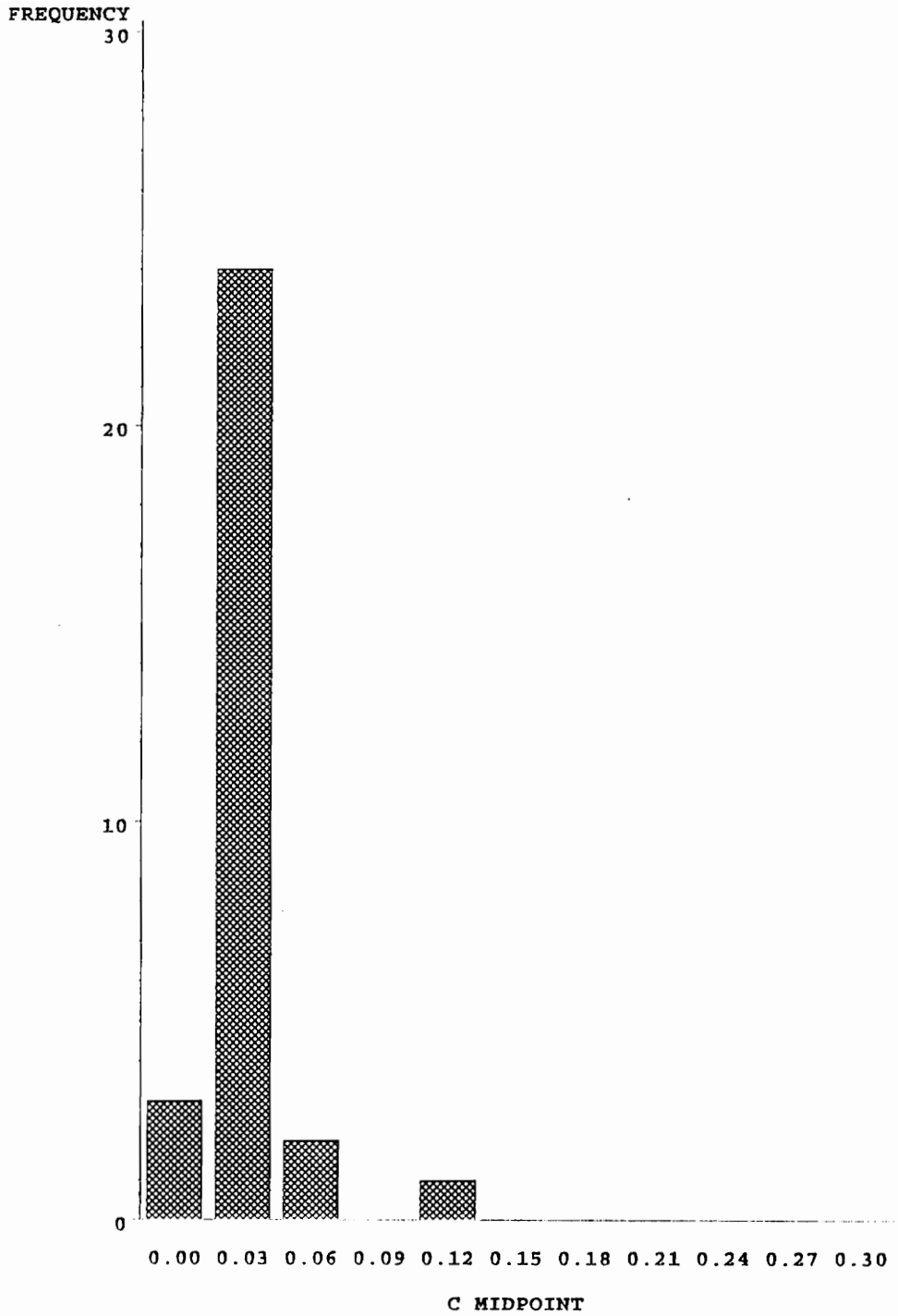

Figure 28: Frequency Distribution of the Cs: J-P New Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items for the J-P scale.

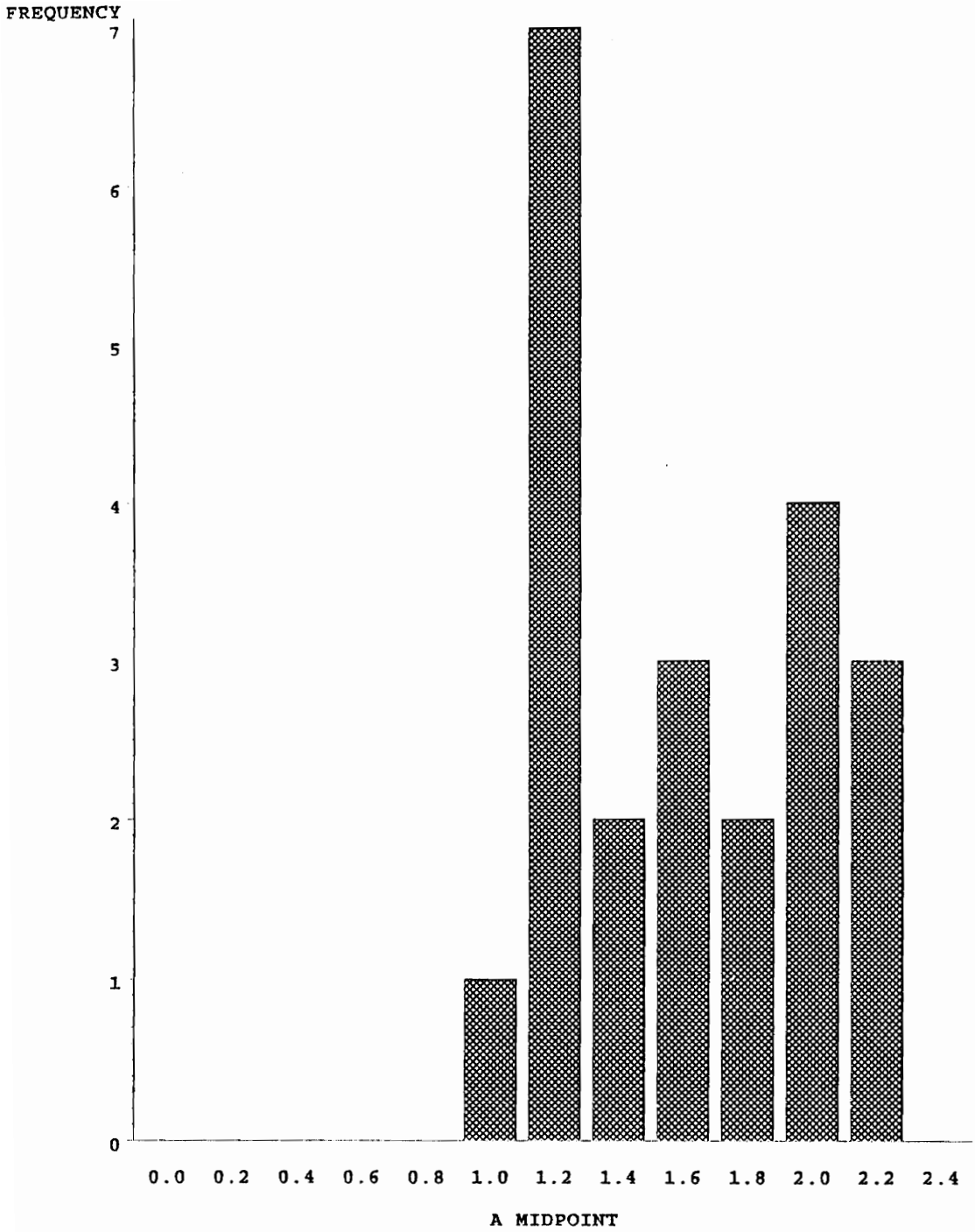

Figure 29: Frequency Distribution of the As: E-I Equal New Items. Frequency distributions of the discrimination (a) parameters for an equal number of new items as those contained in the original MBTI scale for the E-I scale.

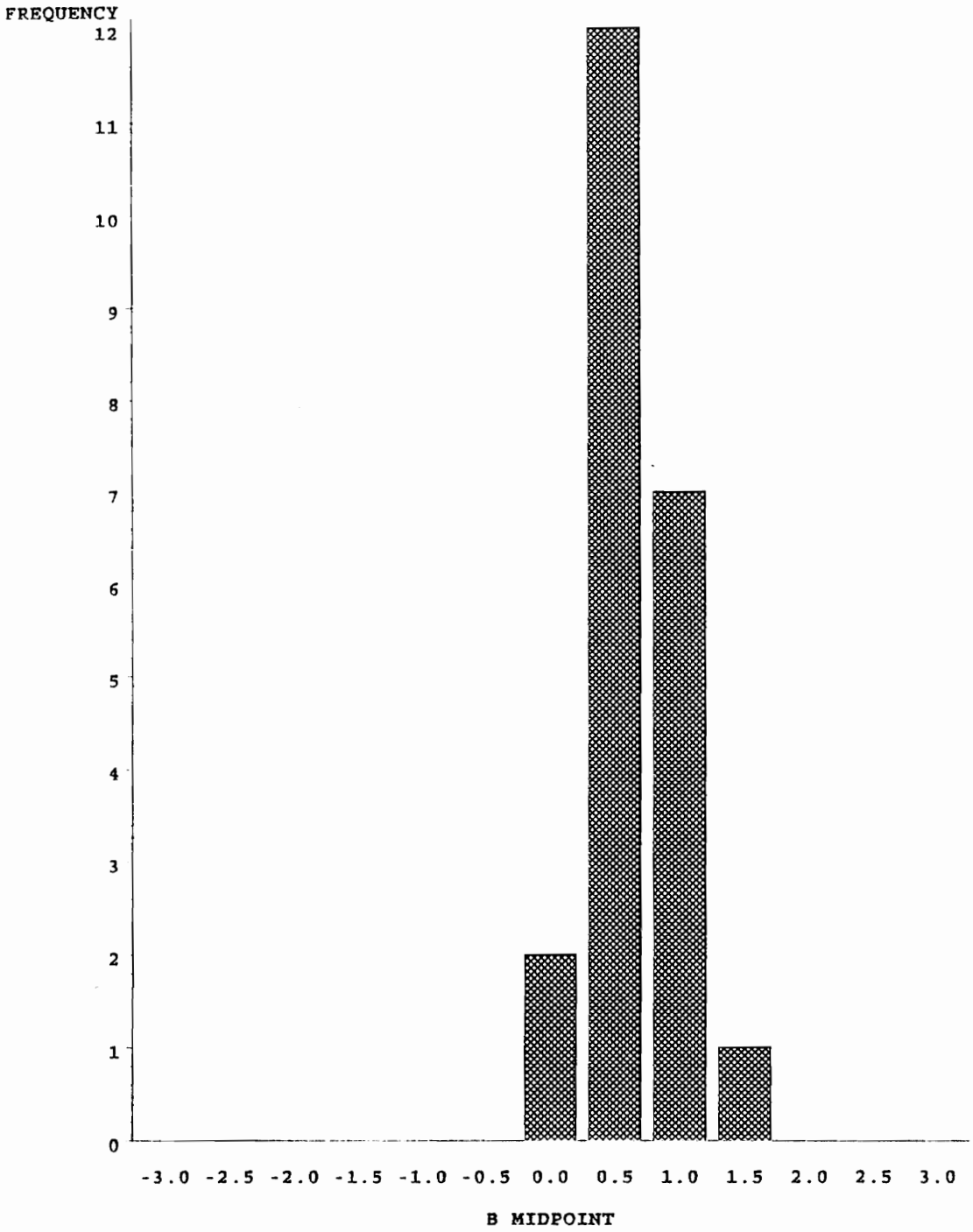

Figure 30: Frequency Distribution of the Bs: E-I Equal New Items. Frequency distributions of the difficulty (b) parameters for an equal number of new items as those contained in the original MBTI scale for the E-I scale.

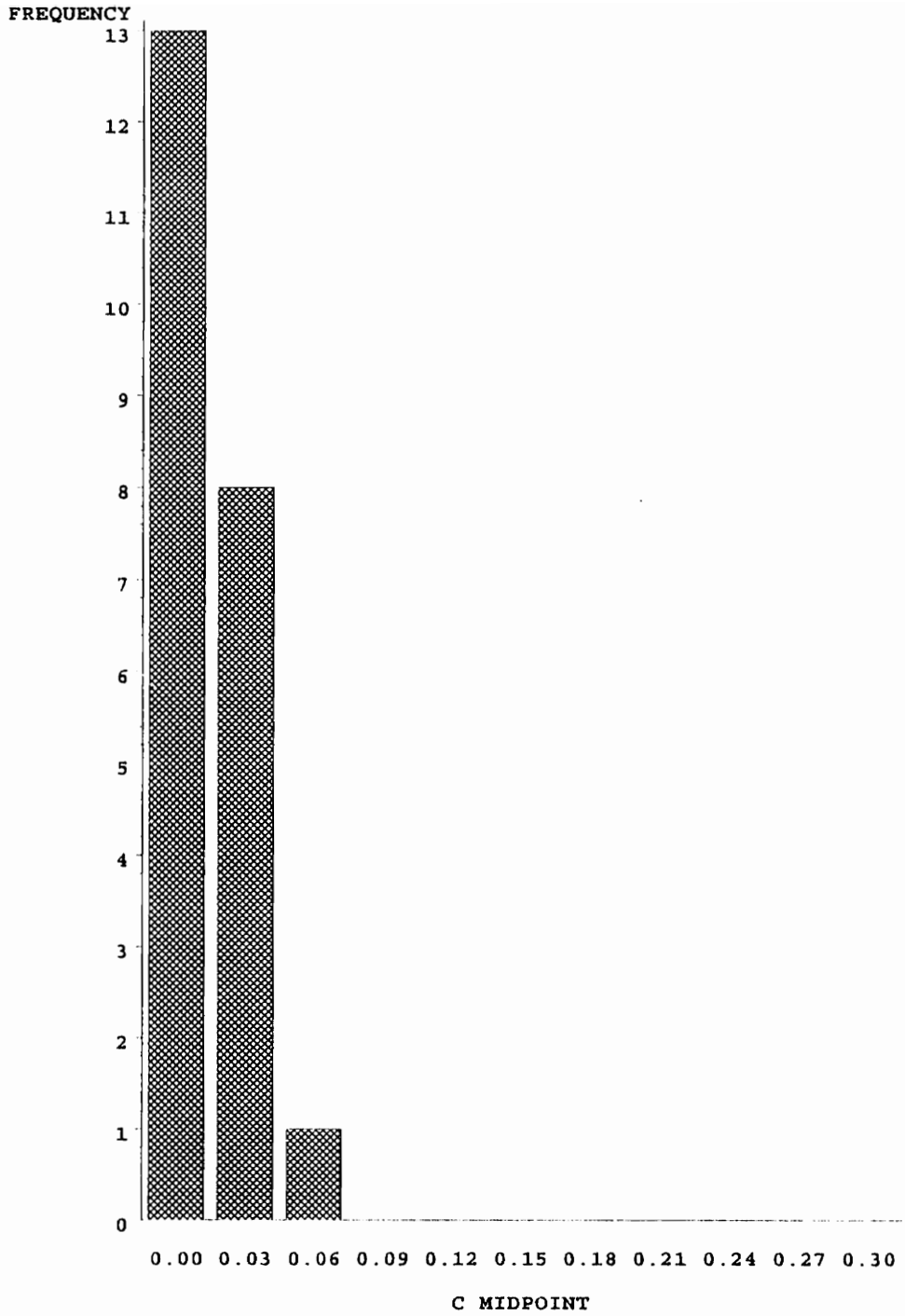

Figure 31: Frequency Distribution of the Cs: E-I Equal New Items. Frequency distributions of the pseudoguessing (c) parameters for an equal number of new items as those contained in the original MBTI scale for the E-I scale.

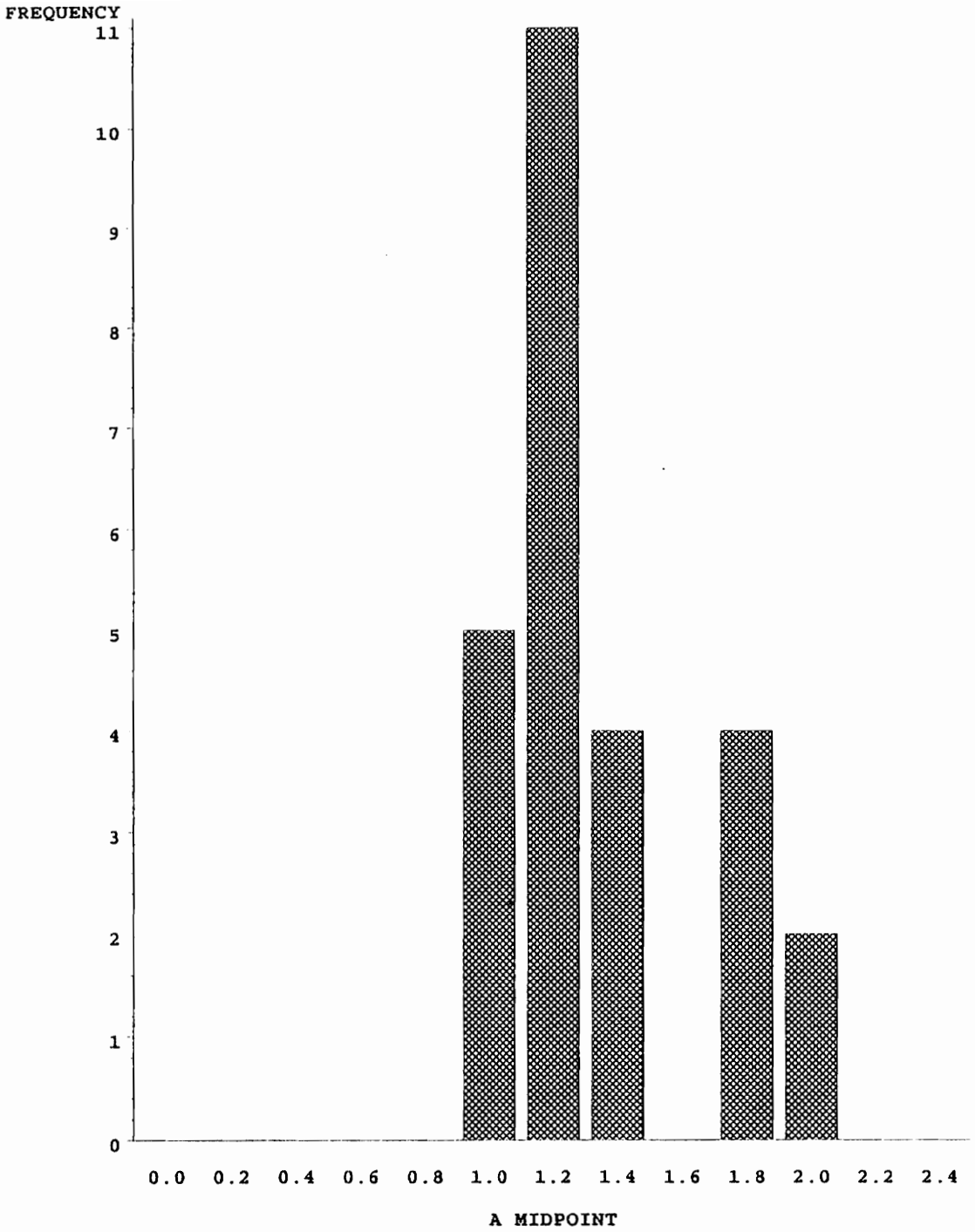

Figure 32: Frequency Distribution of the As: S-N Equal New Items. Frequency distributions of the discrimination (a) parameters for an equal number of new items as those contained in the original MBTI scale for the S-N scale.

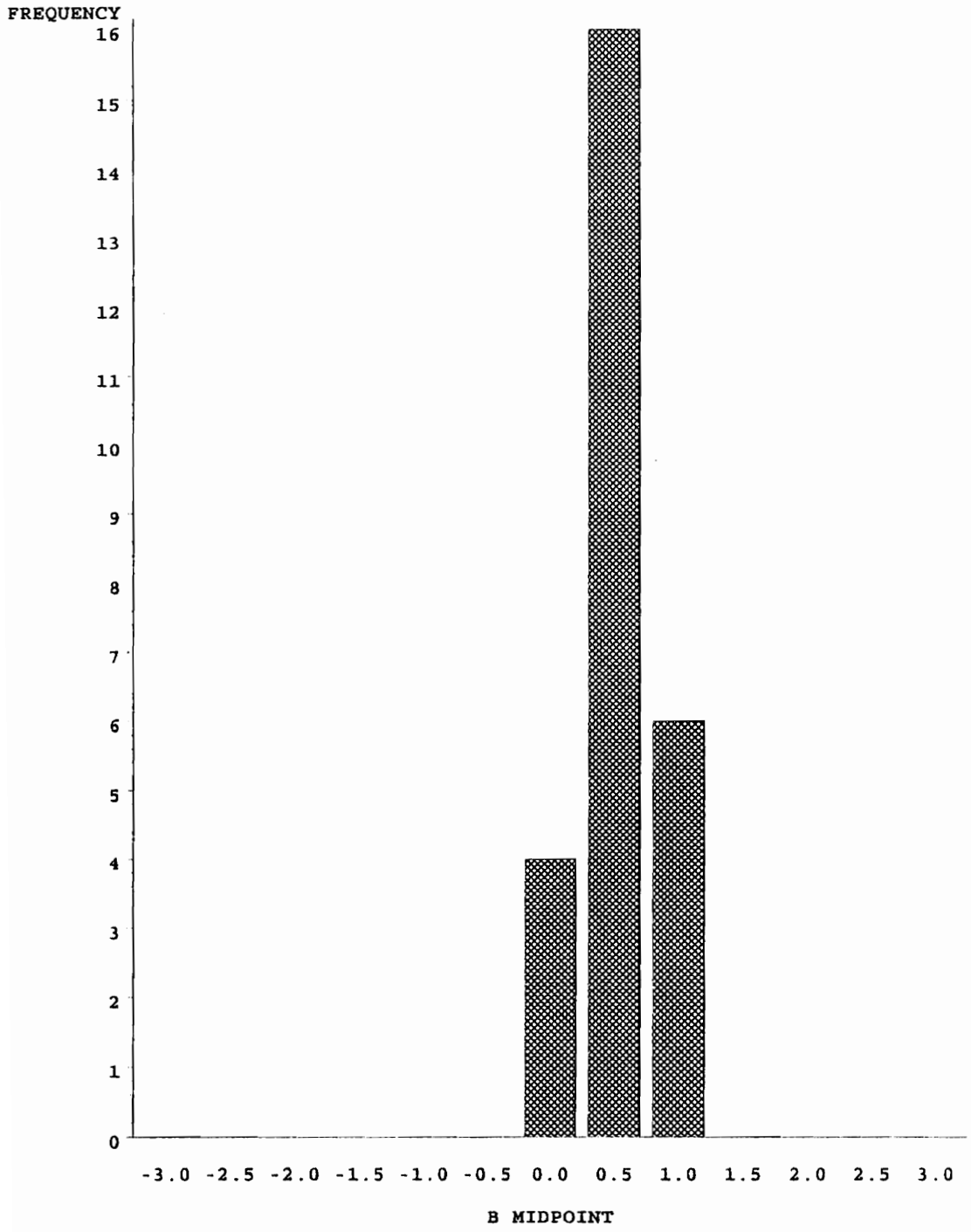

Figure 33: Frequency Distribution of the Bs: S-N Equal New Items. Frequency distributions of the difficulty (b) parameters for an equal number of new items as those contained in the original MBTI scale for the S-N scale.

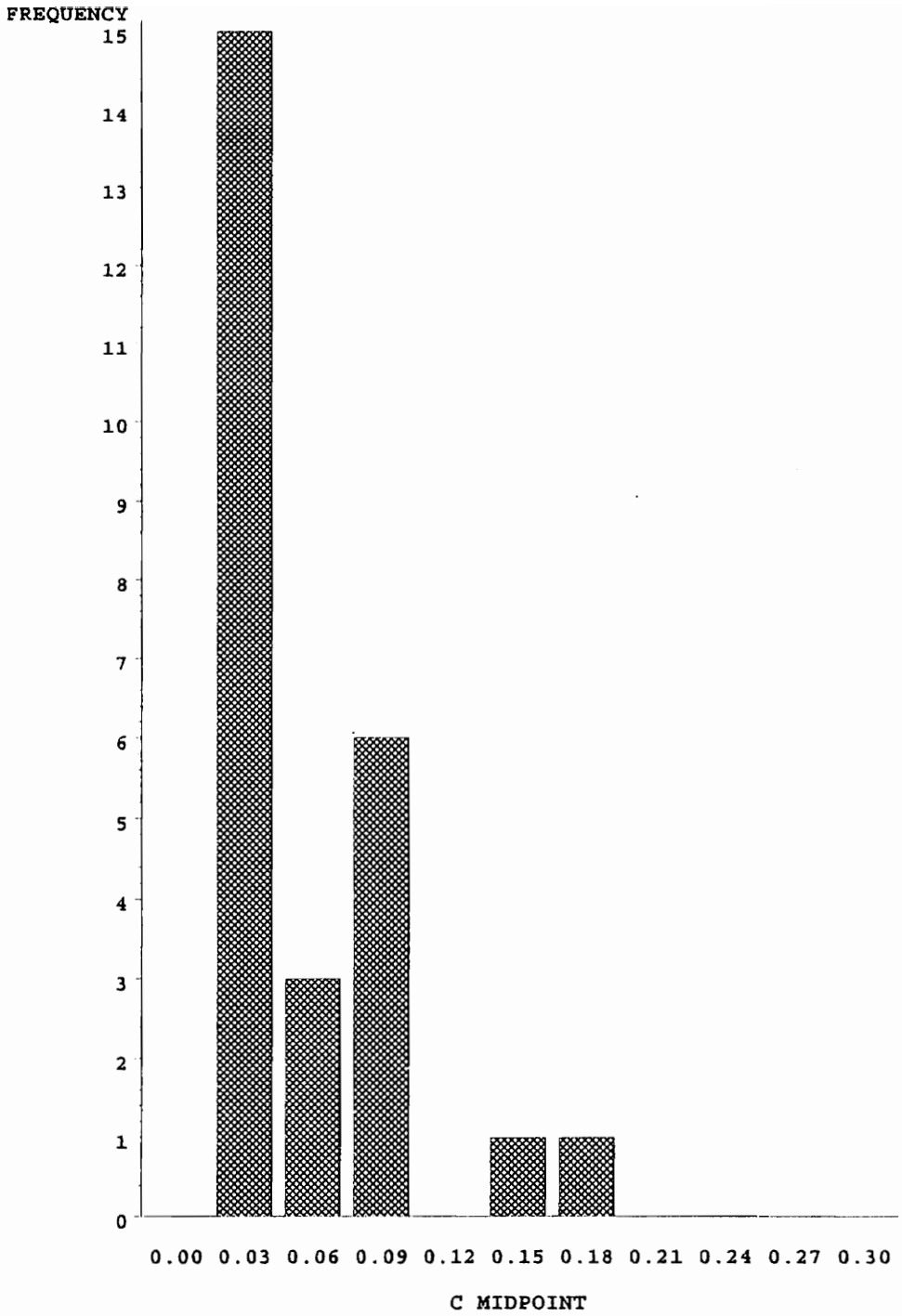

Figure 34: Frequency Distribution of the Cs: S-N Equal New Items. Frequency distributions of the pseudoguessing (c) parameters for an equal number of new items as those contained in the original MBTI scale for the S-N scale.

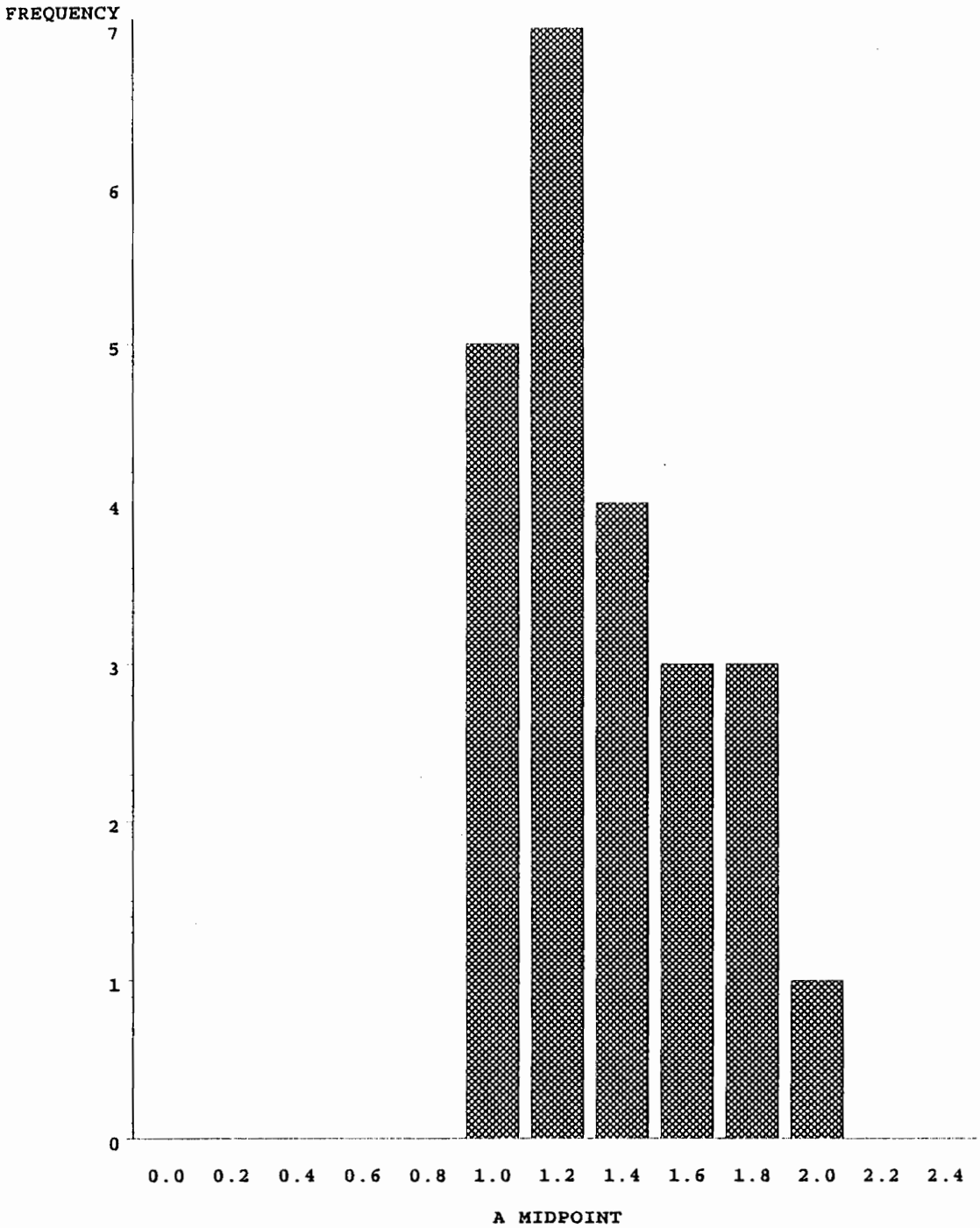

Figure 35: Frequency Distribution of the As: T-F Equal New Items. Frequency distributions of the discrimination (a) parameters for an equal number of new items as those contained in the original MBTI scale for the T-F scale.

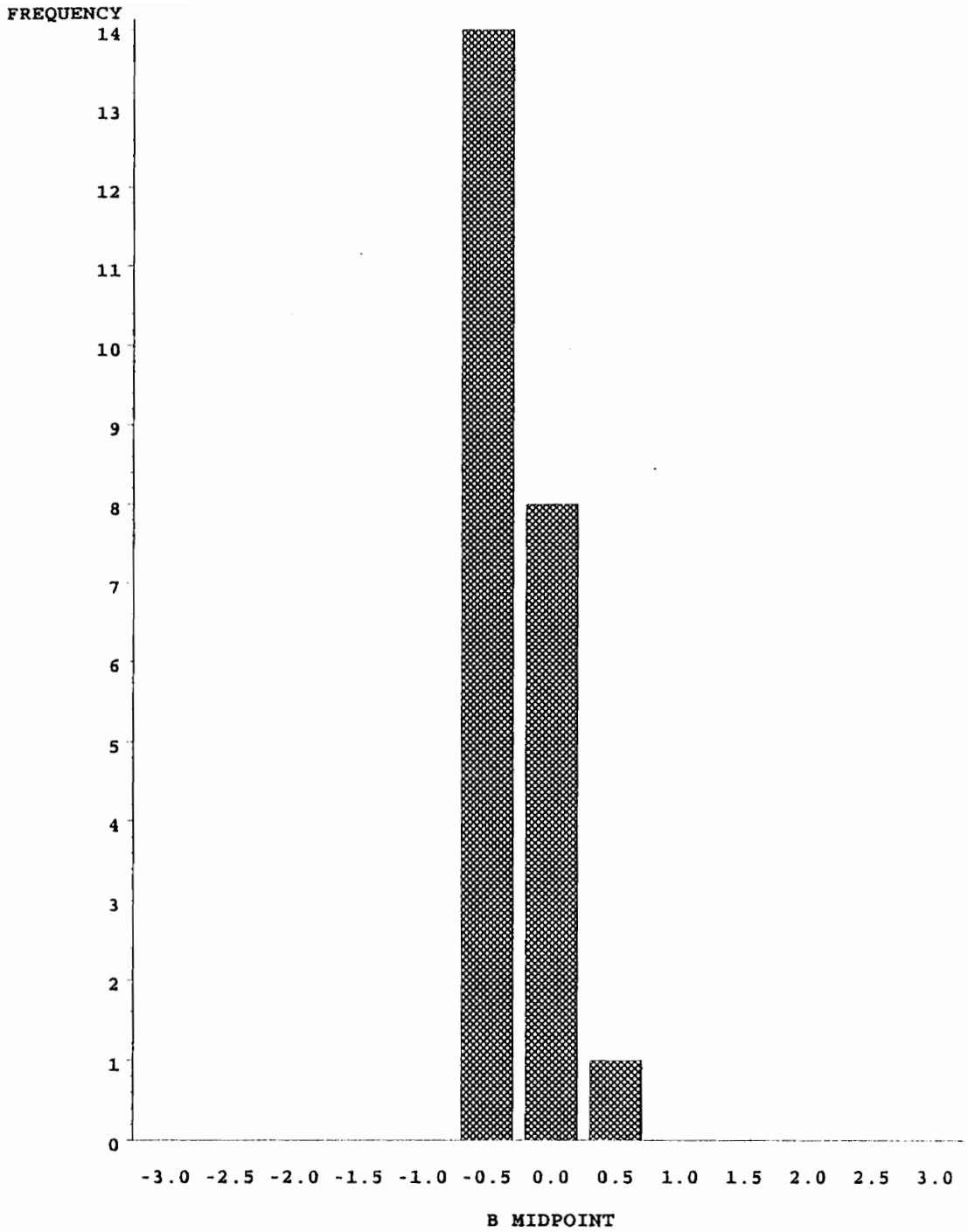

Figure 36: Frequency Distribution of the Bs: T-F Equal New Items. Frequency distributions of the difficulty (b) parameters for an equal number of new items as those contained in the original MBTI scale for the T-F scale.

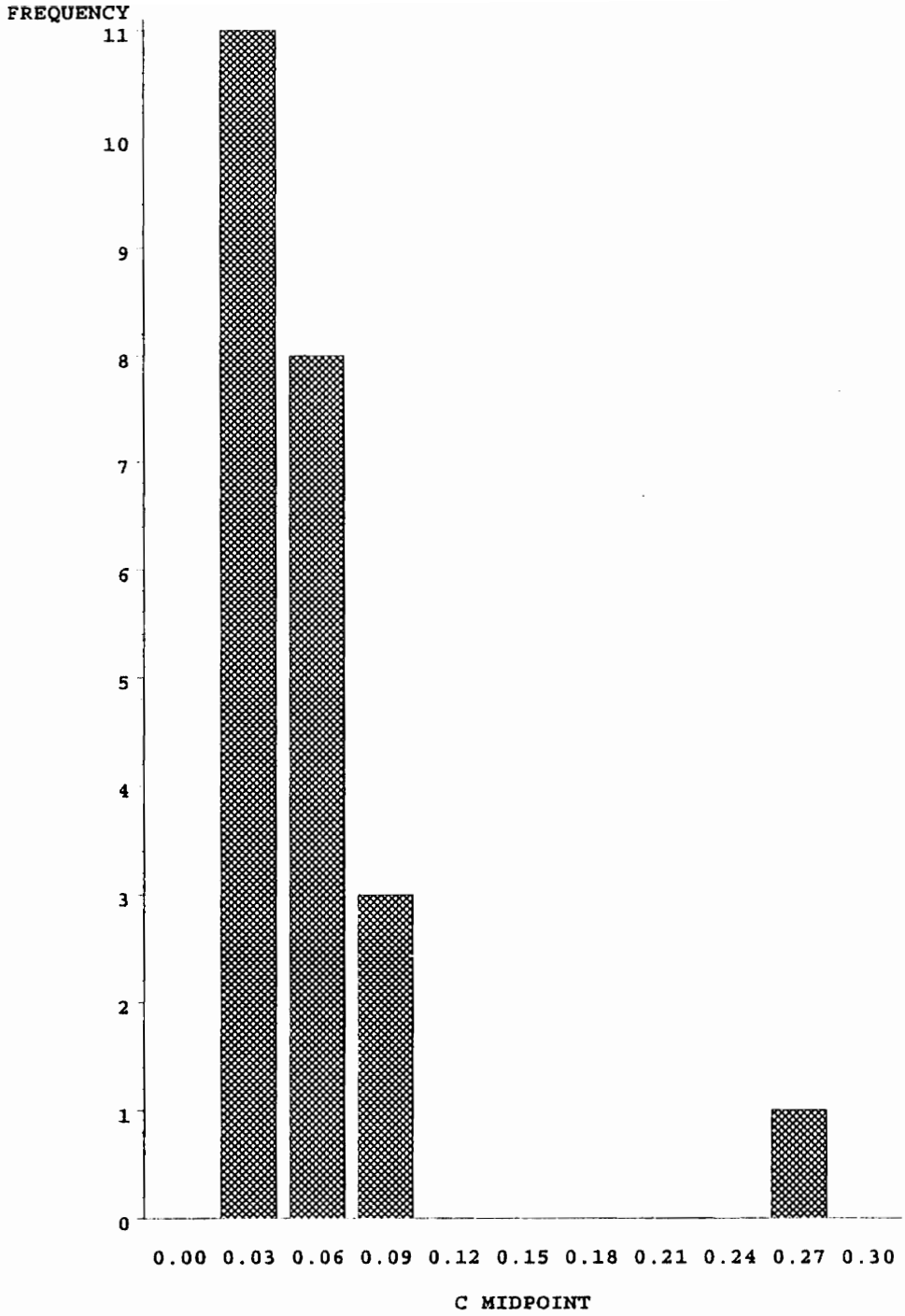

Figure 37: Frequency Distribution of the Cs: T-F Equal New Items. Frequency distributions of the pseudoguessing (c) parameters for an equal number of new items as those contained in the original MBTI scale for the T-F scale.

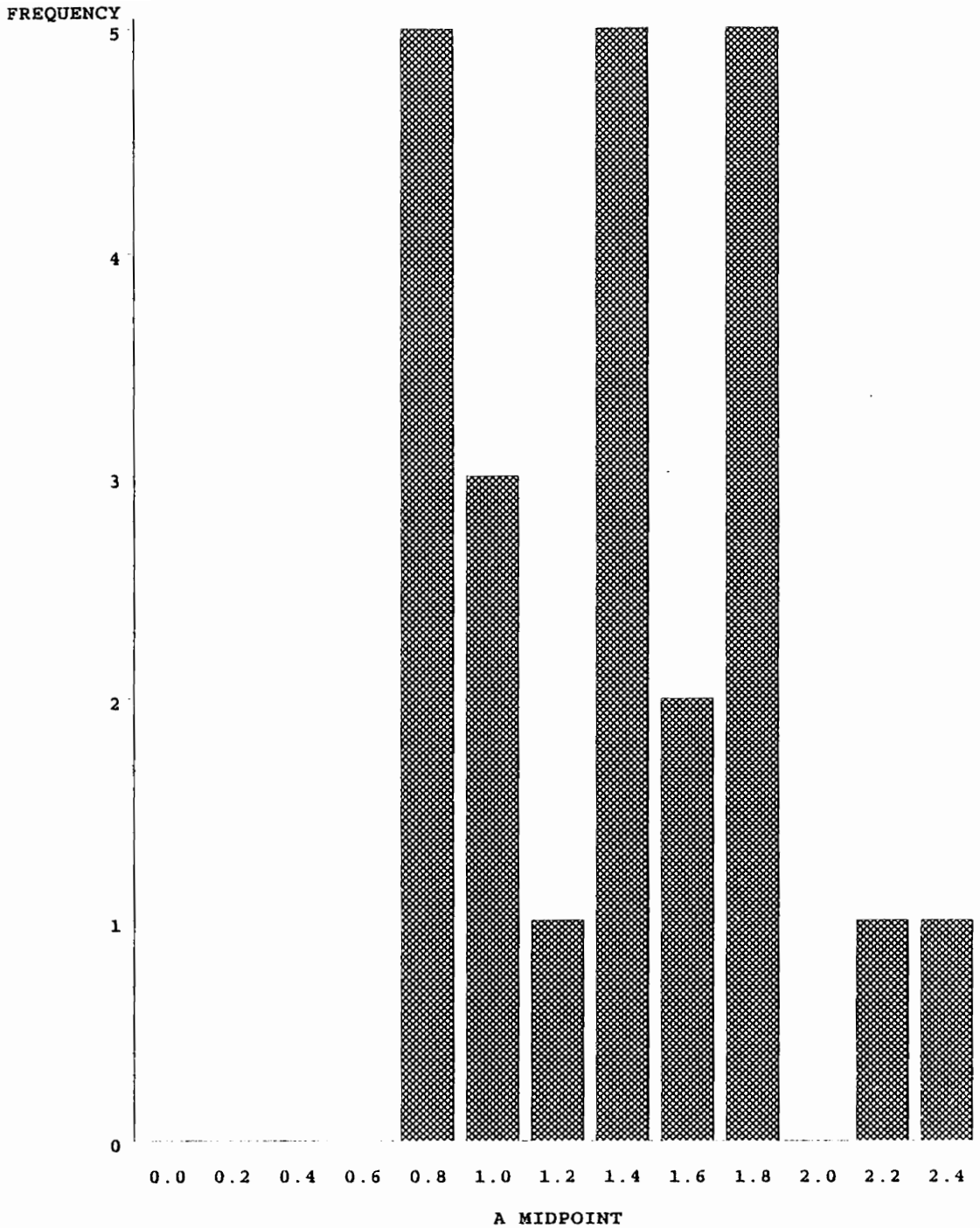

Figure 38: Frequency Distribution of the As: J-P Equal New Items. Frequency distributions of the discrimination (a) parameters for an equal number of new items as those contained in the original MBTI scale for the J-P scale.

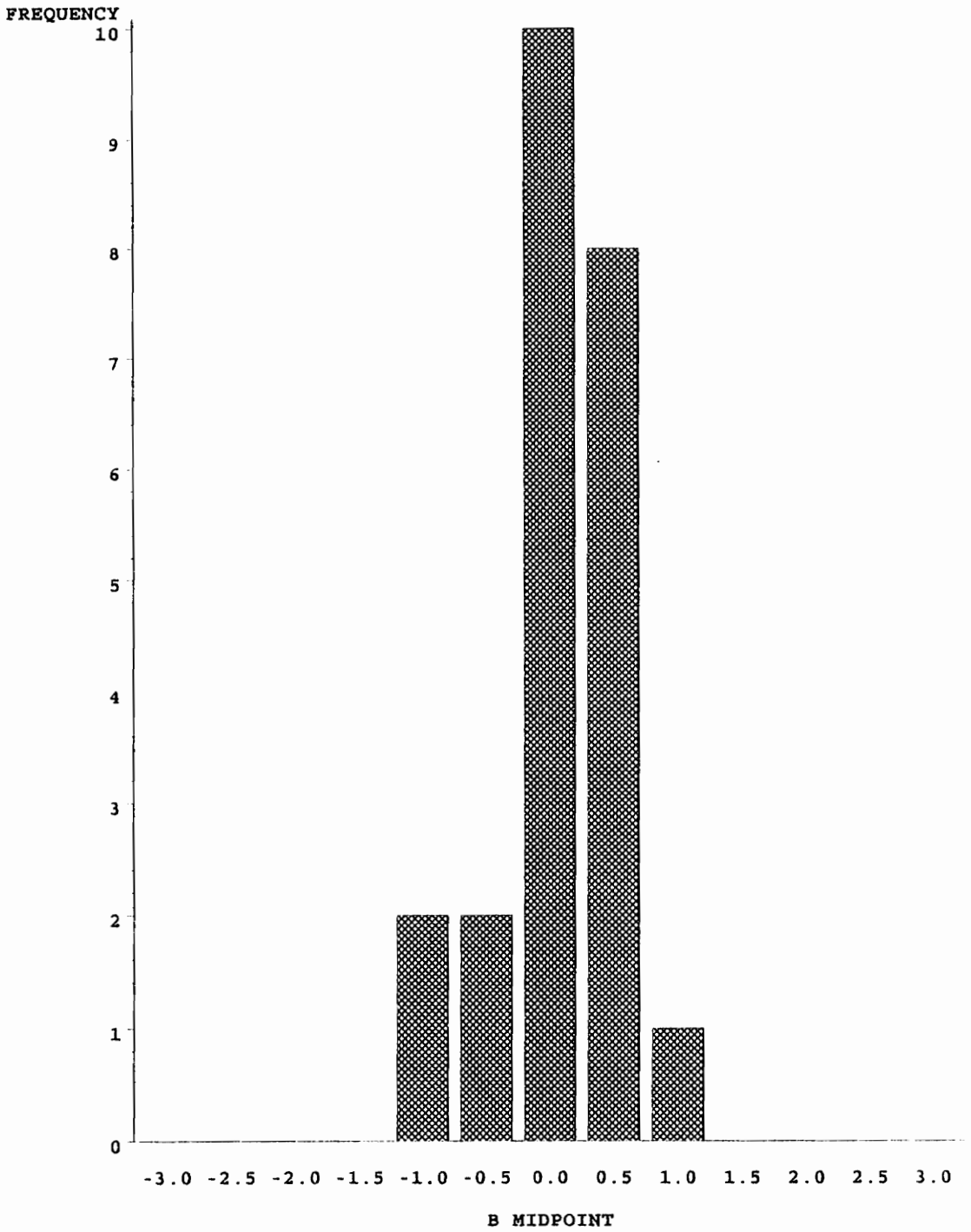

Figure 39: Frequency Distribution of the Bs: J-P Equal New Items. Frequency distributions of the difficulty (b) parameters for an equal number of new items as those contained in the original MBTI scale for the J-P scale.

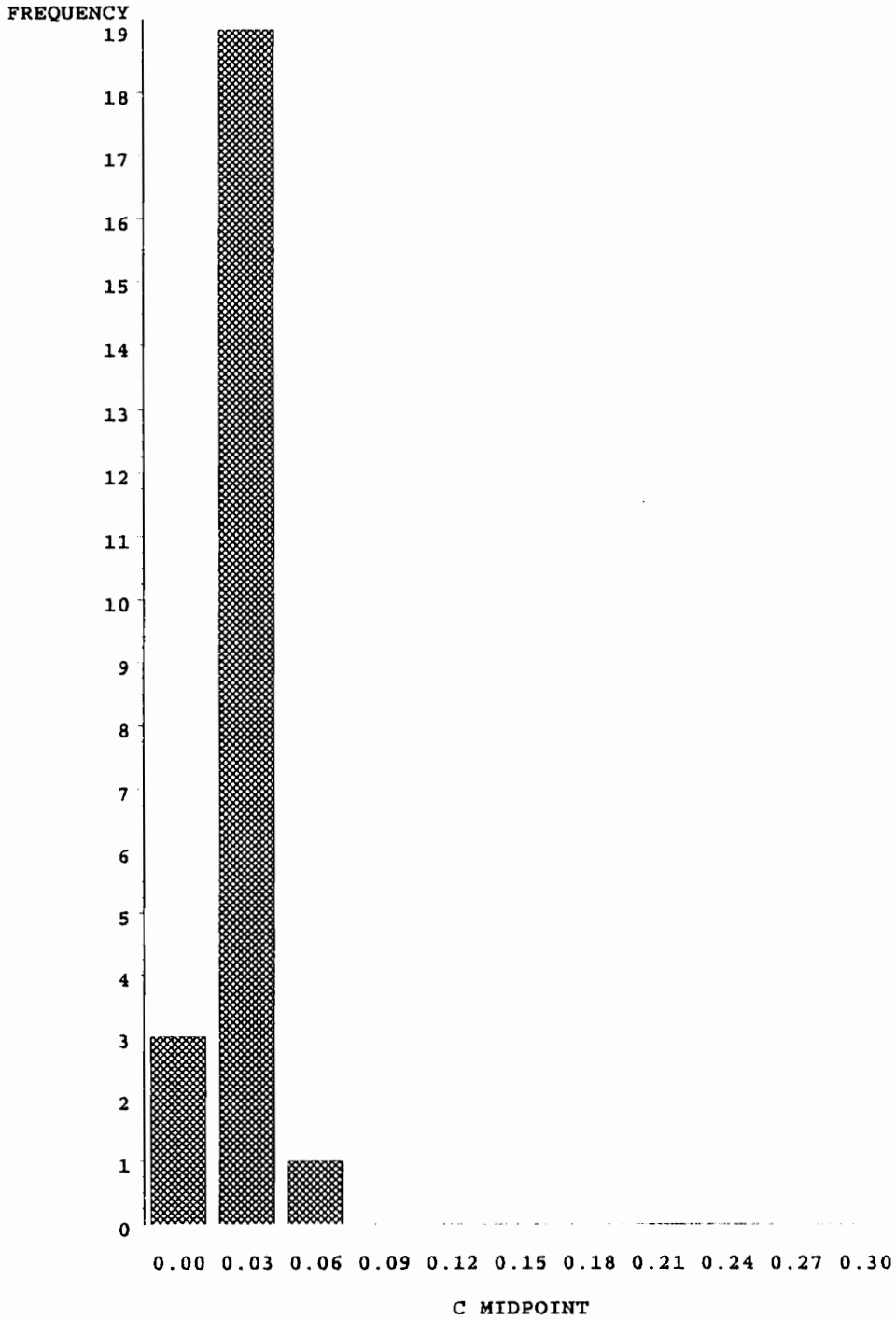

Figure 40: Frequency Distribution of the Cs: J-P Equal New Items. Frequency distributions of the pseudoguessing (c) parameters for an equal number of new items as those contained in the original MBTI scale for the J-P scale.

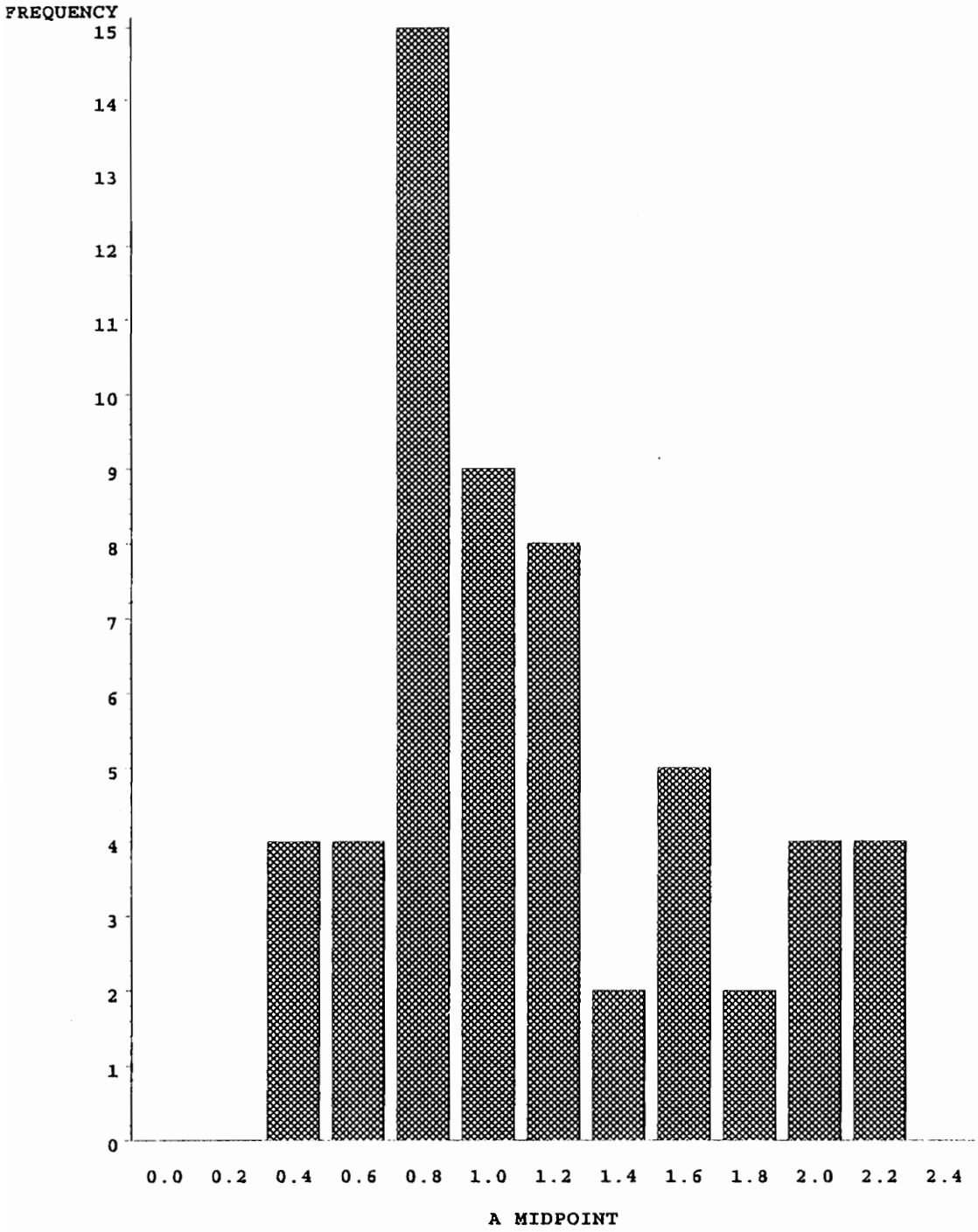

Figure 41: Frequency Distribution of the As: E-I Combined Items. Frequency distributions of the discrimination (a) parameters for all the the new items combined with the original MBTI items for the E-I scale.

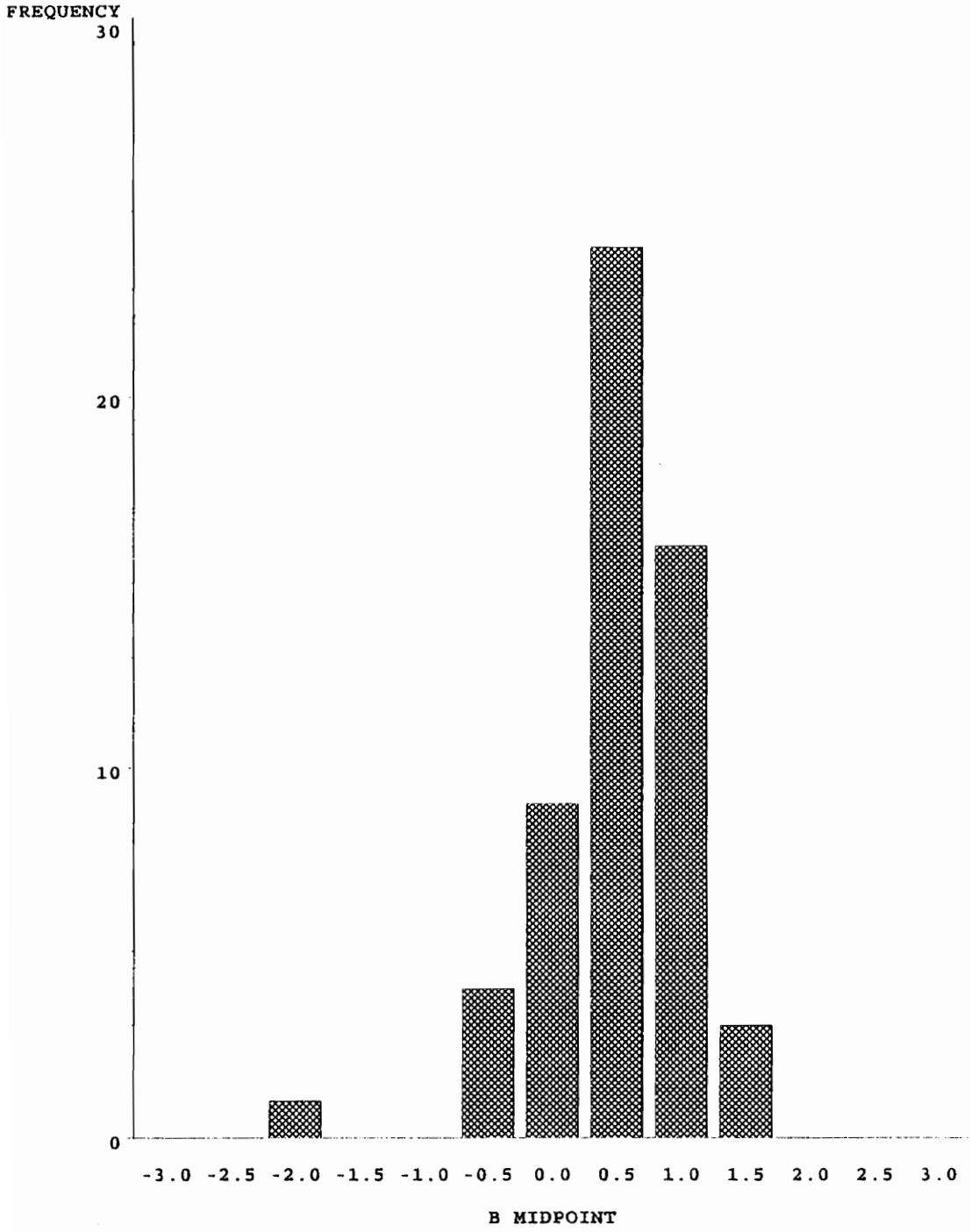

Figure 42: Frequency Distribution of the Bs: E-I Combined Items. Frequency distributions of the difficulty (b) parameters for all the new items combined with the original MBTI items for the E-I scale.

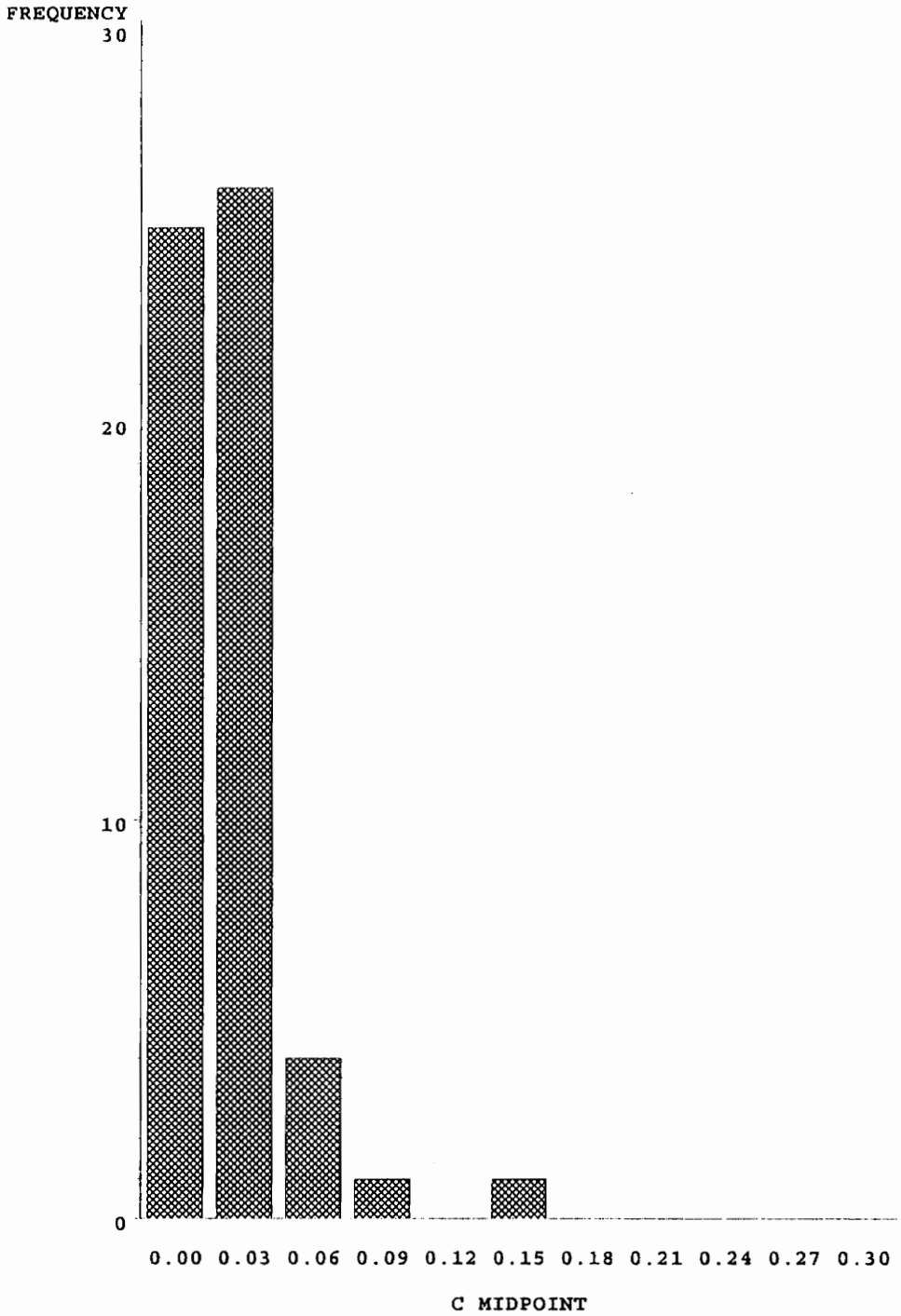

Figure 43: Frequency Distribution of the Cs: E-I Combined Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items combined with the original MBTI items for the E-I scale.

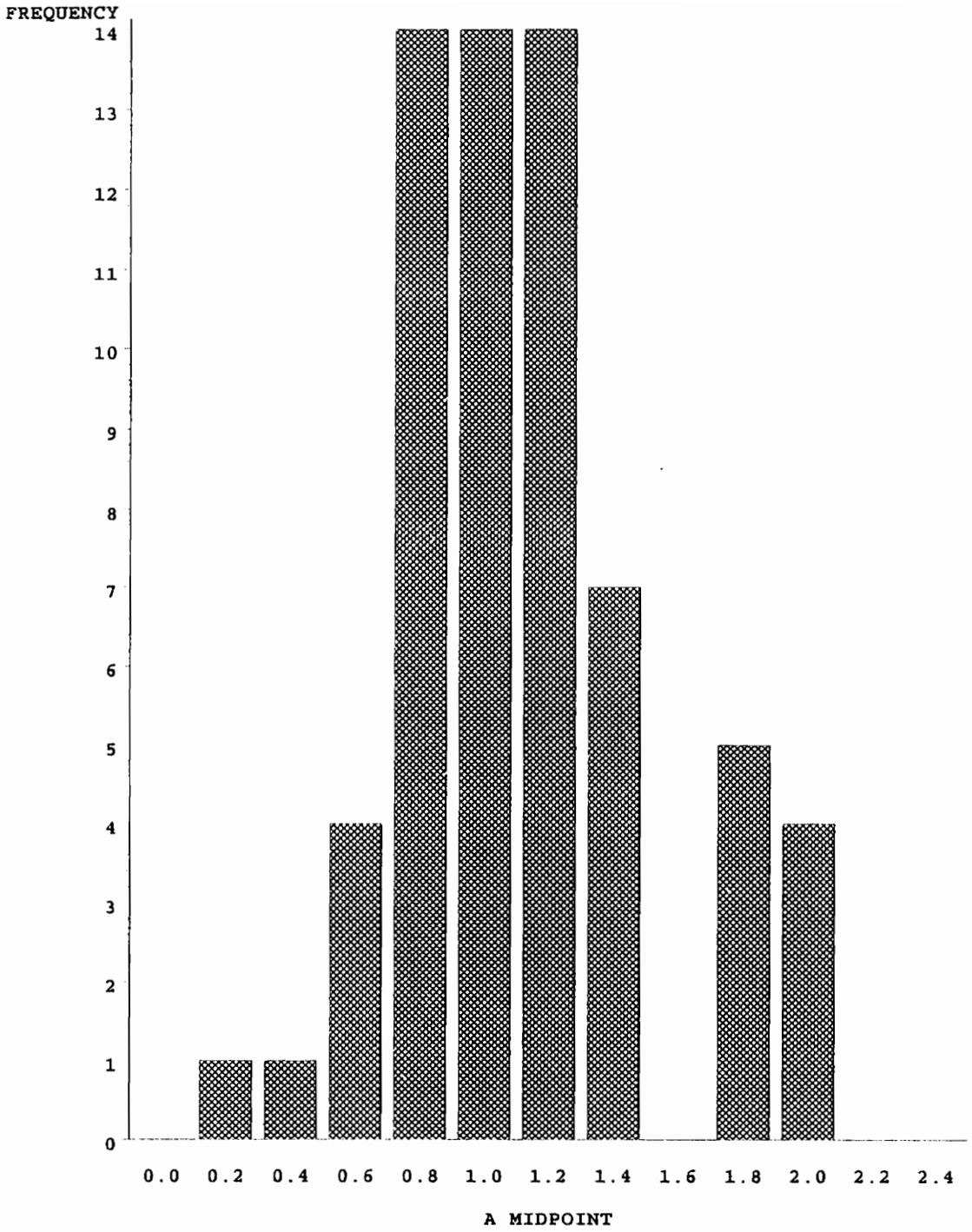

Figure 44: Frequency Distribution of the As: S-N Combined Items. Frequency distributions of the discrimination (a) parameters for all the new items combined with the original MBTI items for the S-N scale.

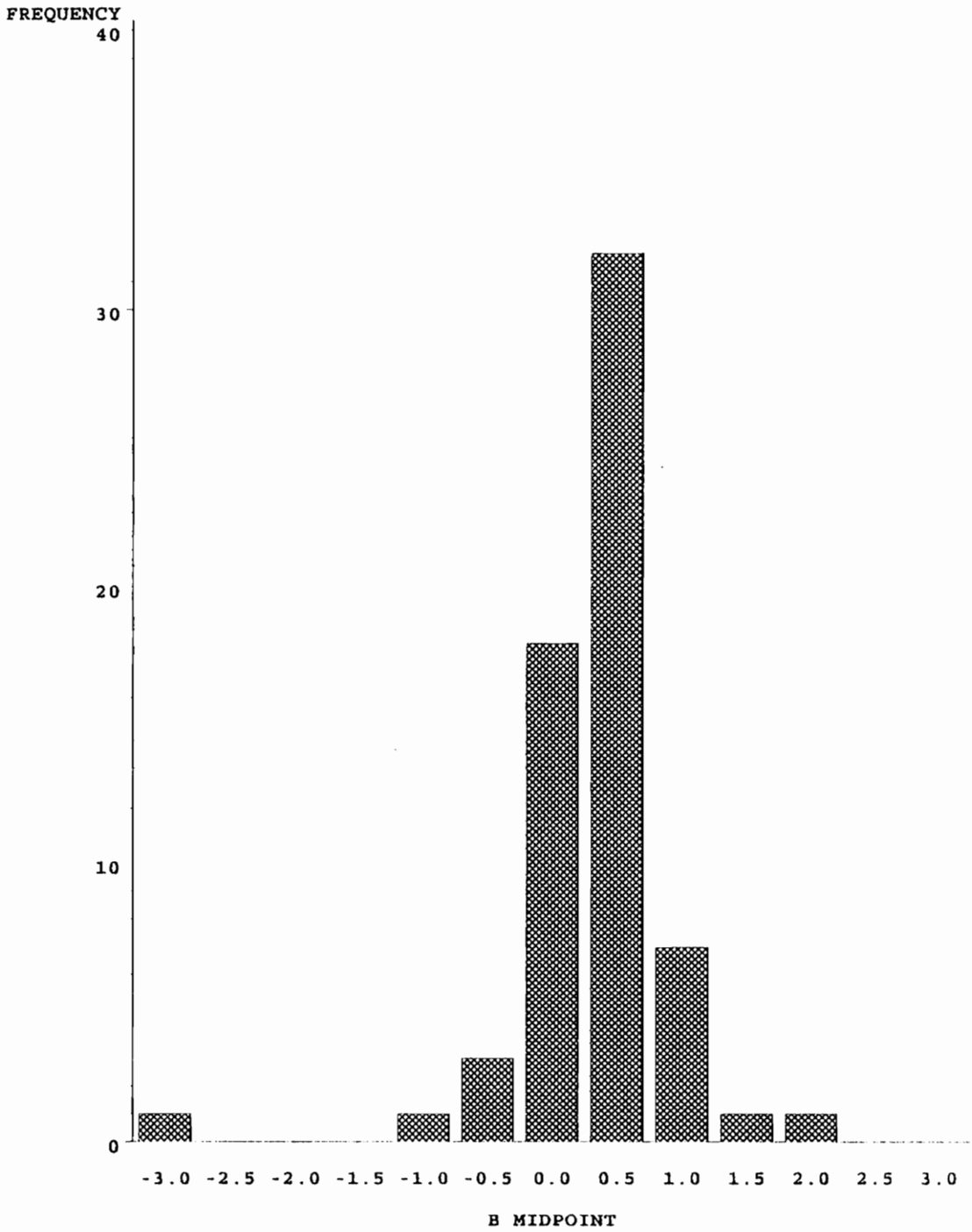

Figure 45: Frequency Distribution of the Bs: S-N Combined Items. Frequency distributions of the difficulty (b) parameters for all the new items combined with the original MBTI items for the S-N scale.

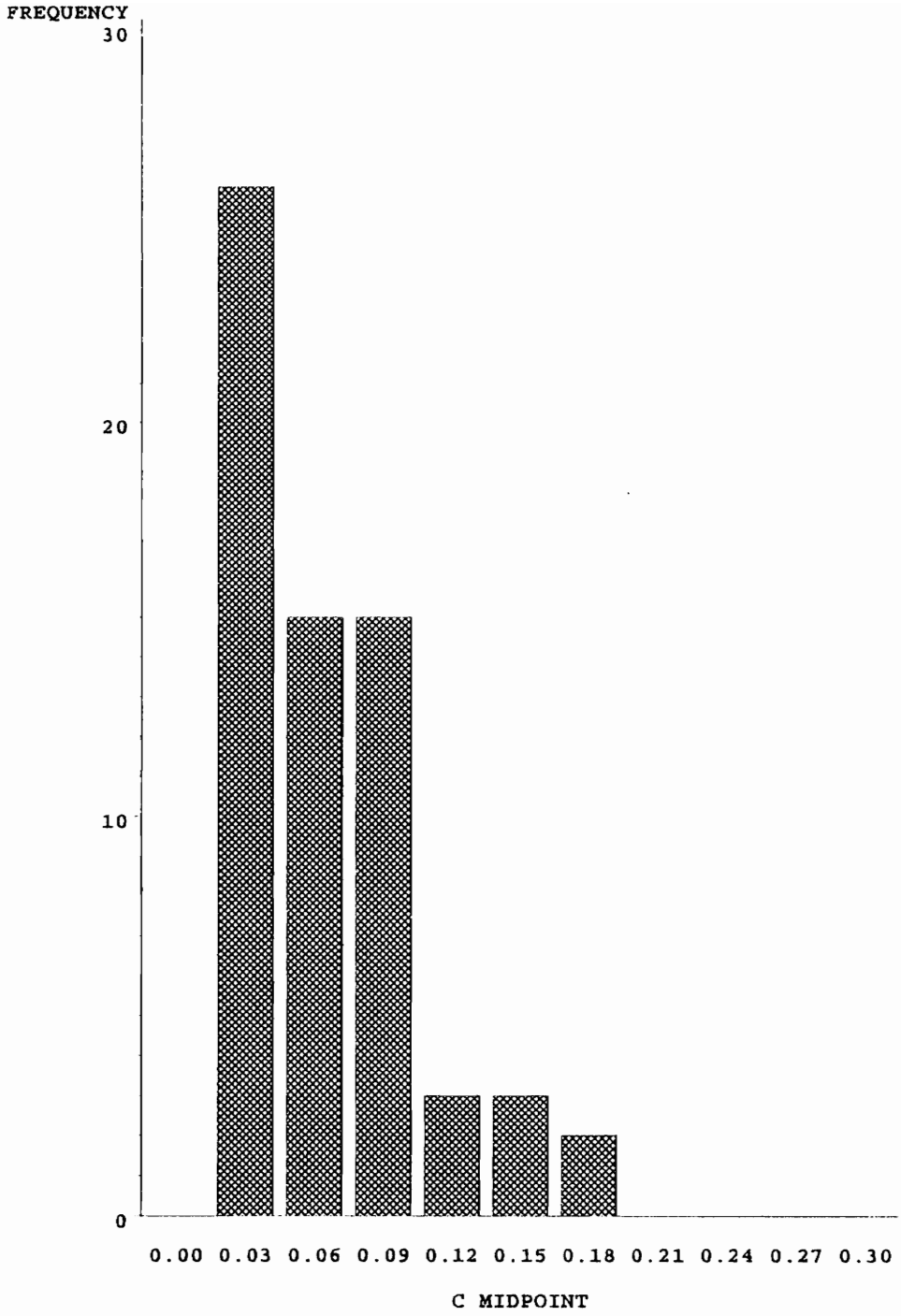

Figure 46: Frequency Distribution of the Cs: S-N Combined Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items combined with the original MBTI items for the S-N scale.

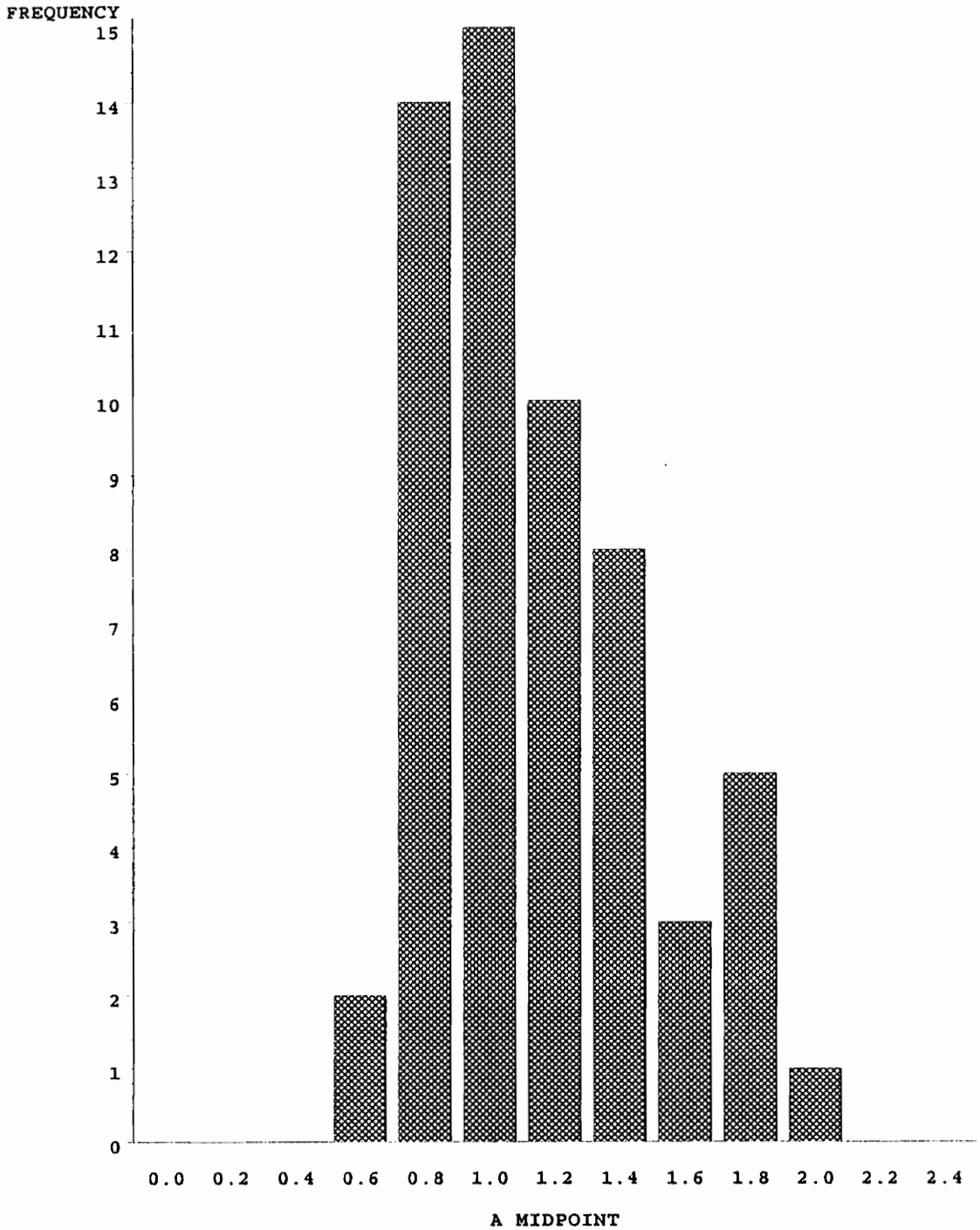

Figure 47 : Frequency Distribution of the As: T-F Combined Items. Frequency distributions of the discrimination (a) parameters for all the new items combined with the original MBTI items for the T-F scale.

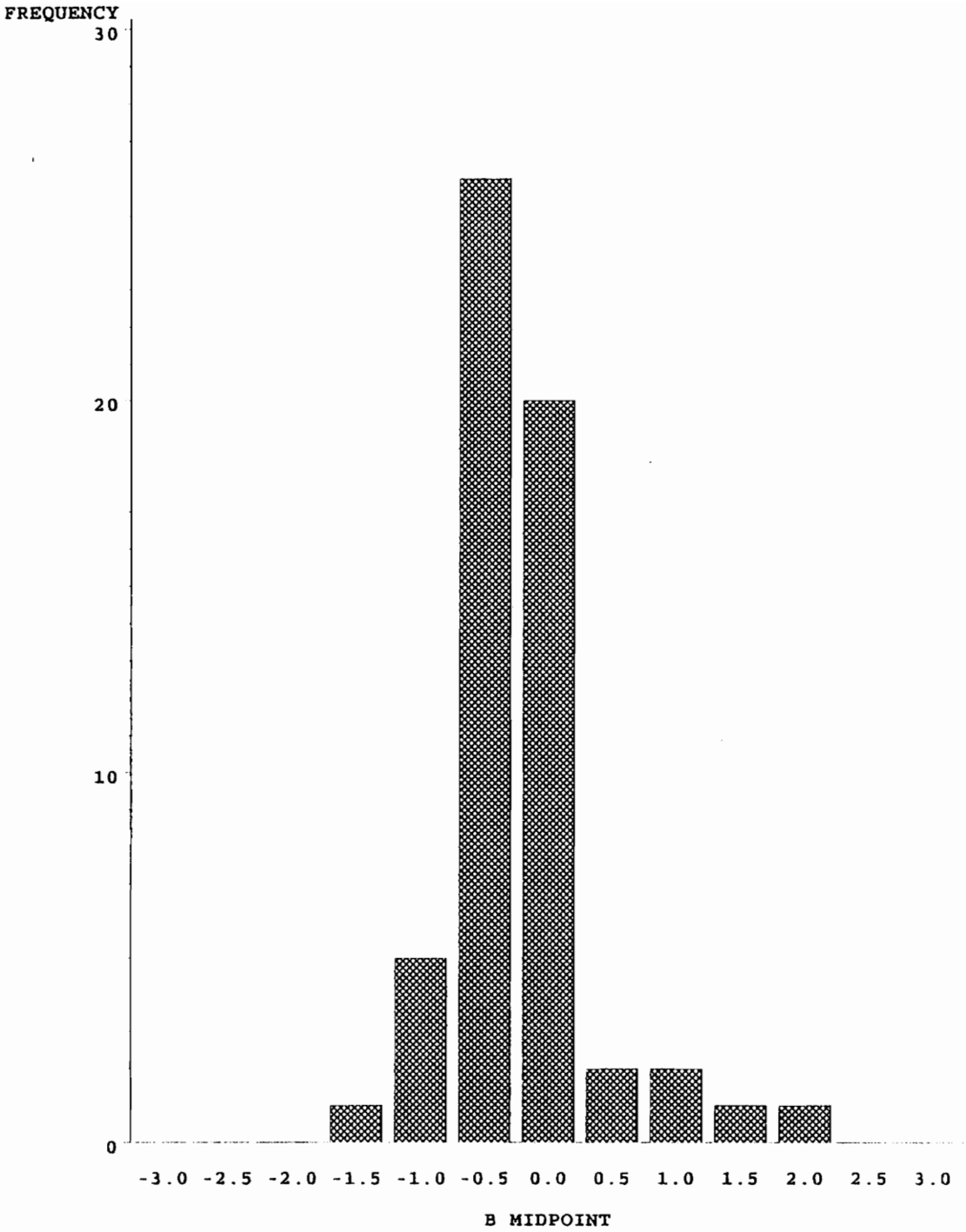

Figure 48: Frequency Distribution of the Bs: T-F Combined Items. . Frequency distributions of the difficulty (b) parameters for all the new items combined with the original MBTI items for the T-F scale.

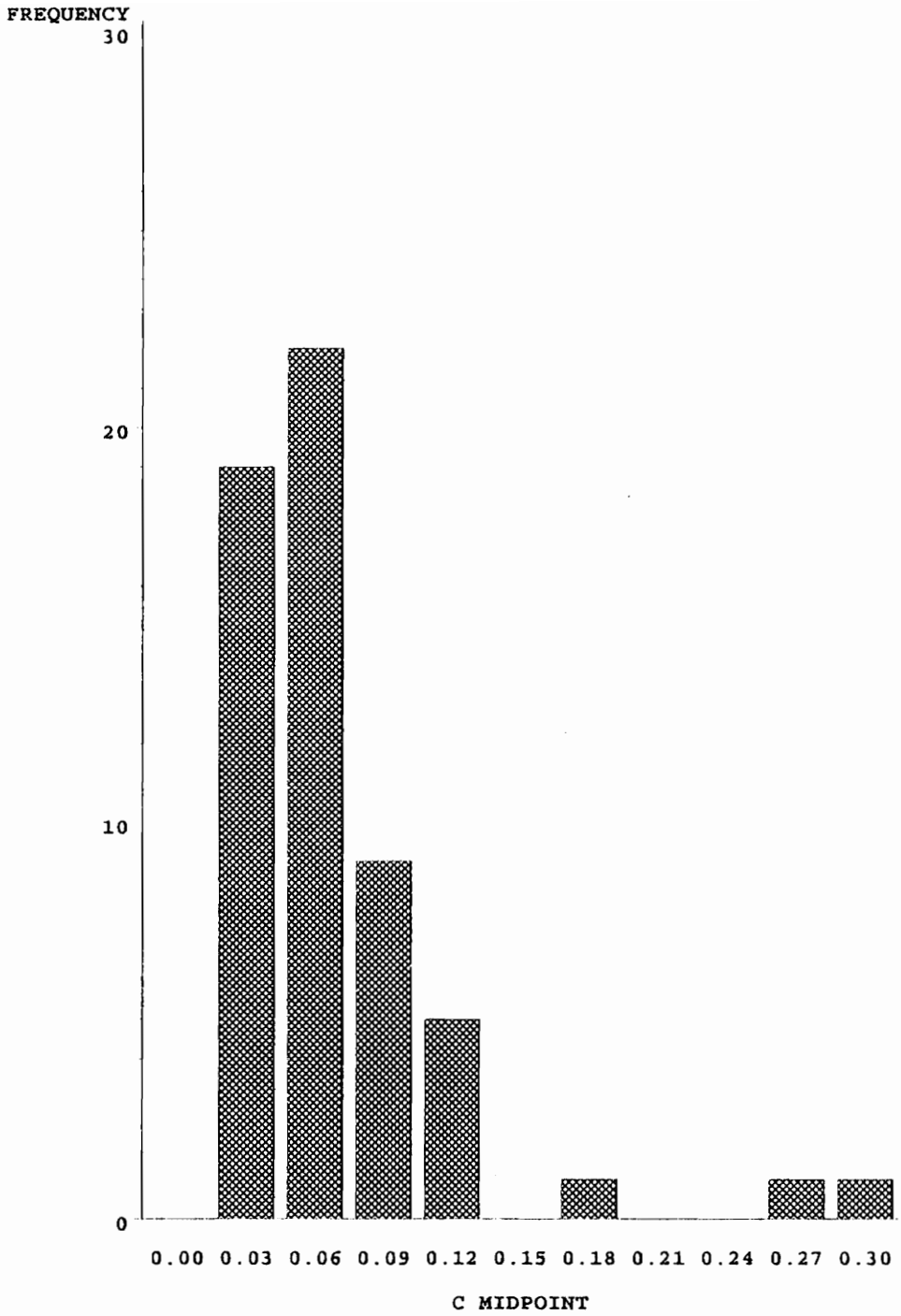

Figure 49: Frequency Distribution of the Cs: T-F Combined Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items combined with the original MBTI items for the T-F scale.

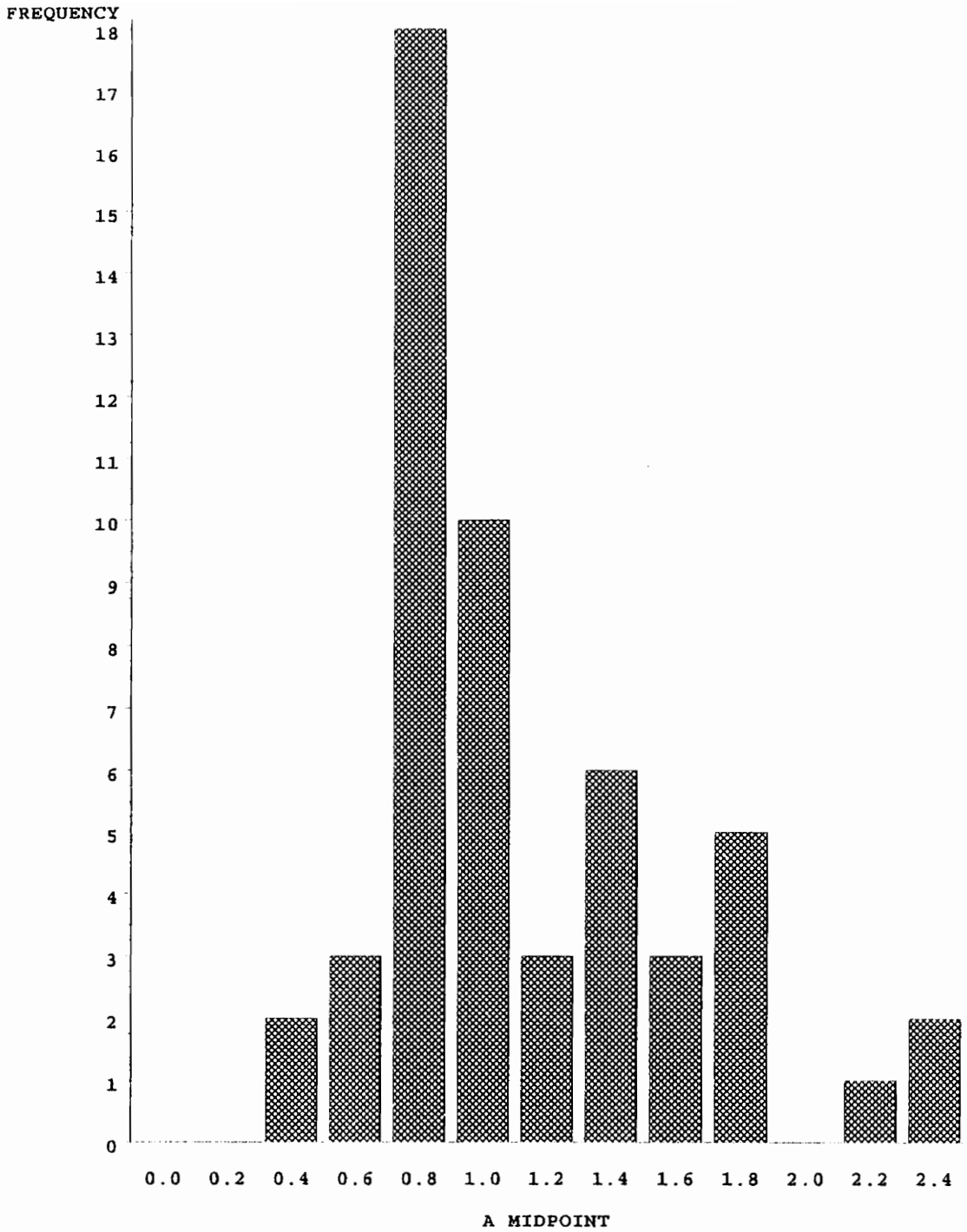

Figure 50: Frequency Distribution of the As: J-P Combined Items. Frequency distributions of the discrimination (a) parameters for all the new items combined with the original MBTI items for the J-P scale.

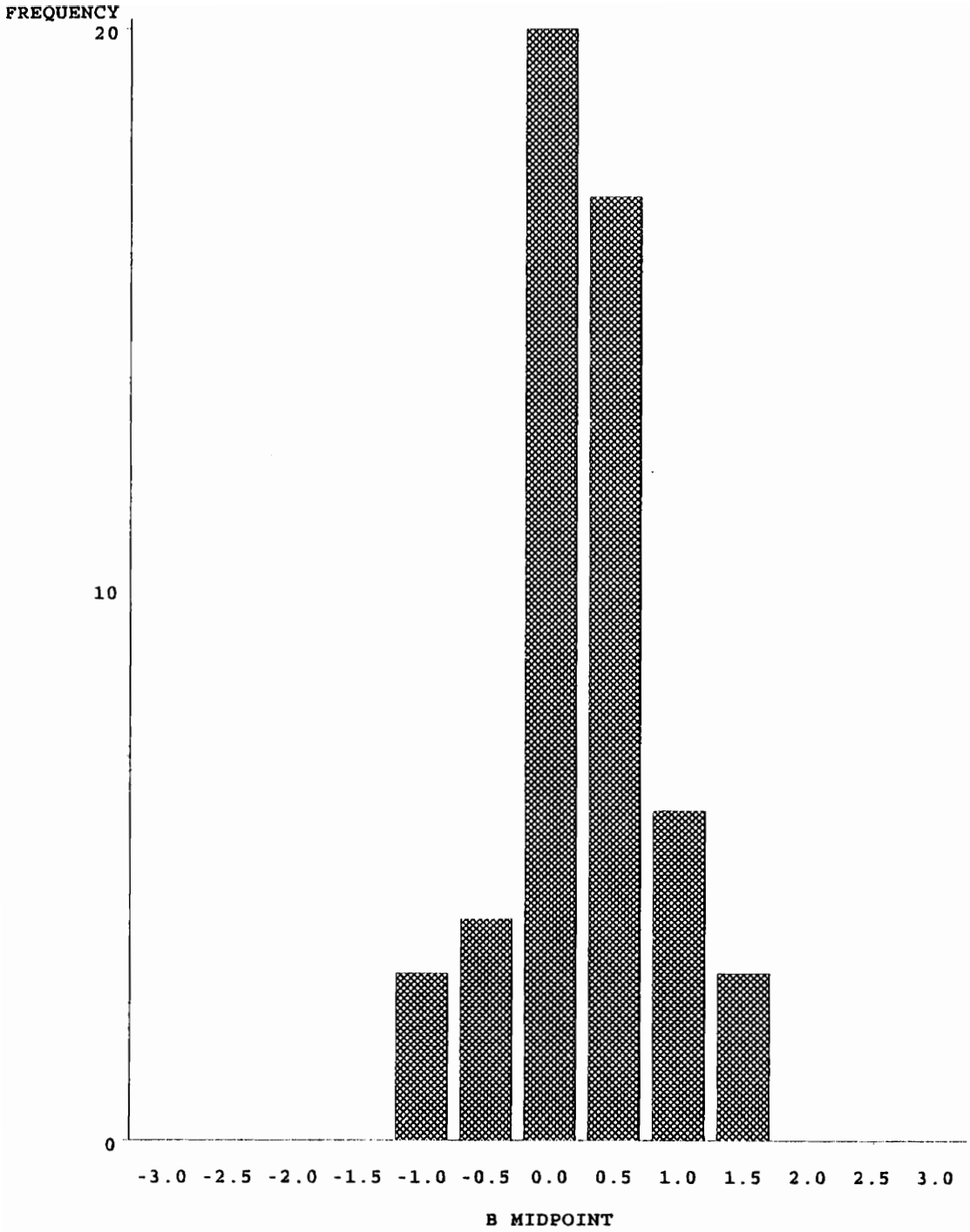

Figure 51: Frequency Distribution of the Bs: J-P Combined Items. Frequency distributions of the difficulty (b) parameters for all the new items combined with the original MBTI items for the J-P scale.

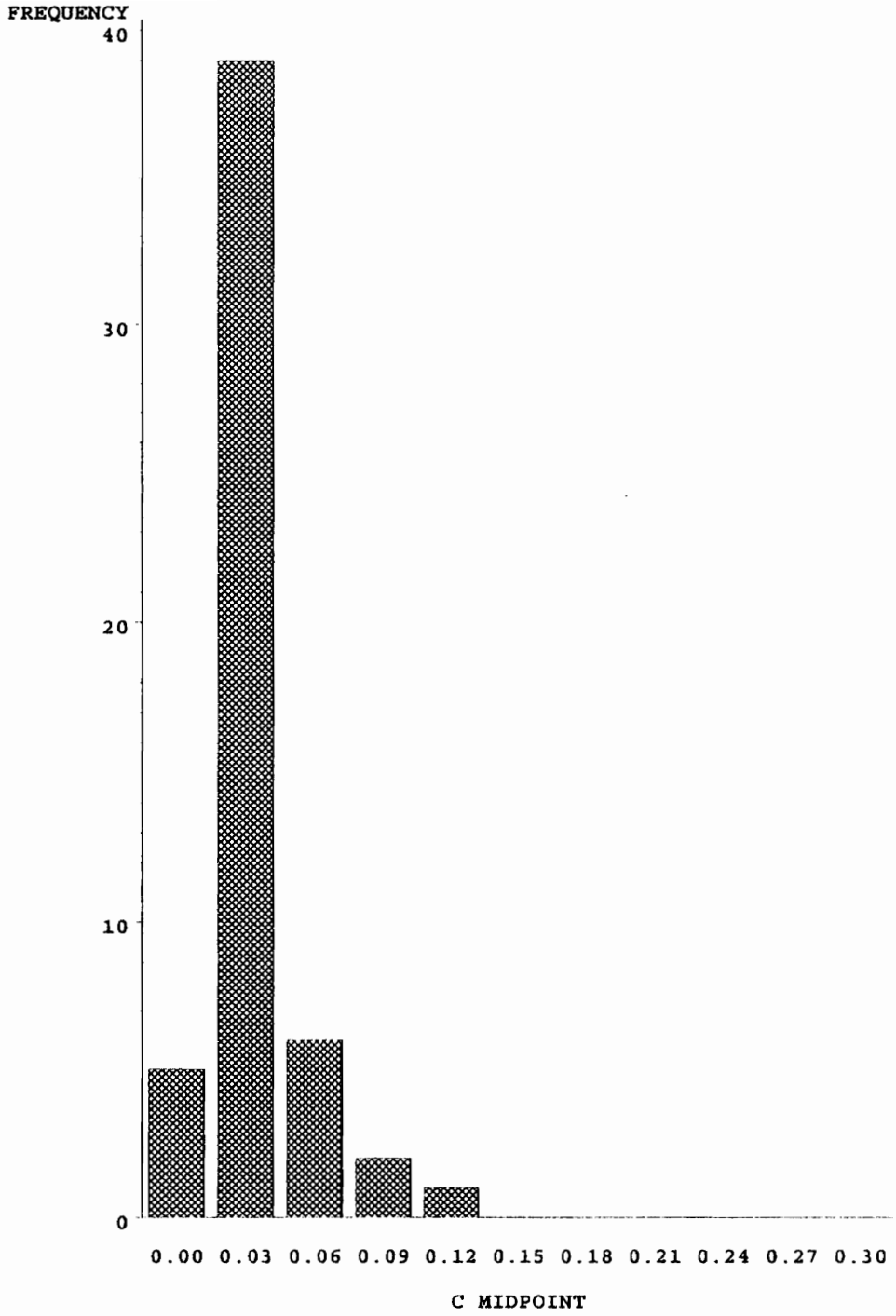

Figure 52: Frequency Distribution of the Cs: J-P Combined Items. Frequency distributions of the pseudoguessing (c) parameters for all the new items combined with the original MBTI items for the J-P scale.

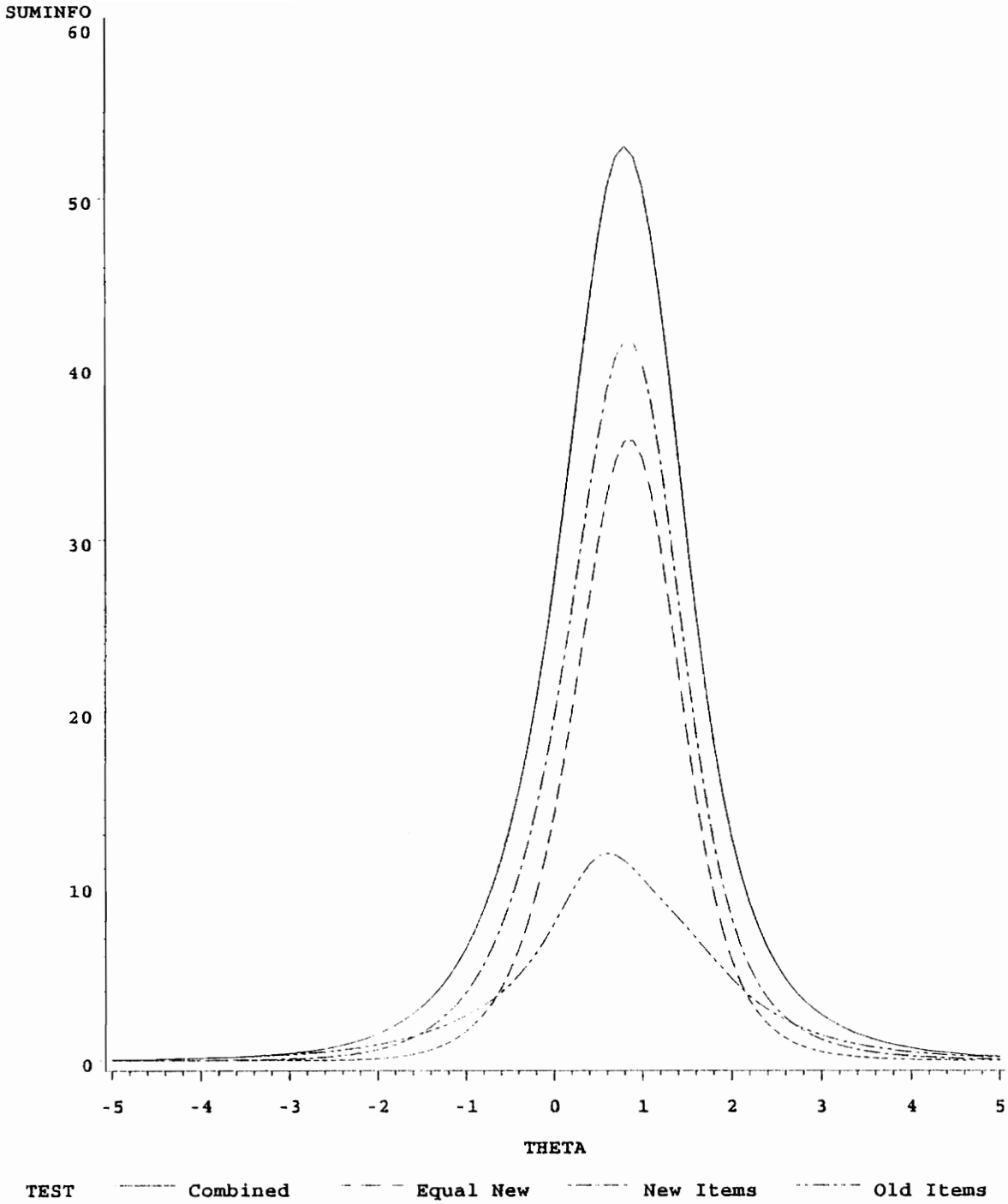

Figure 53: Test Information Functions for the E-I Scale. Test information functions for four different item pools within the E/I scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

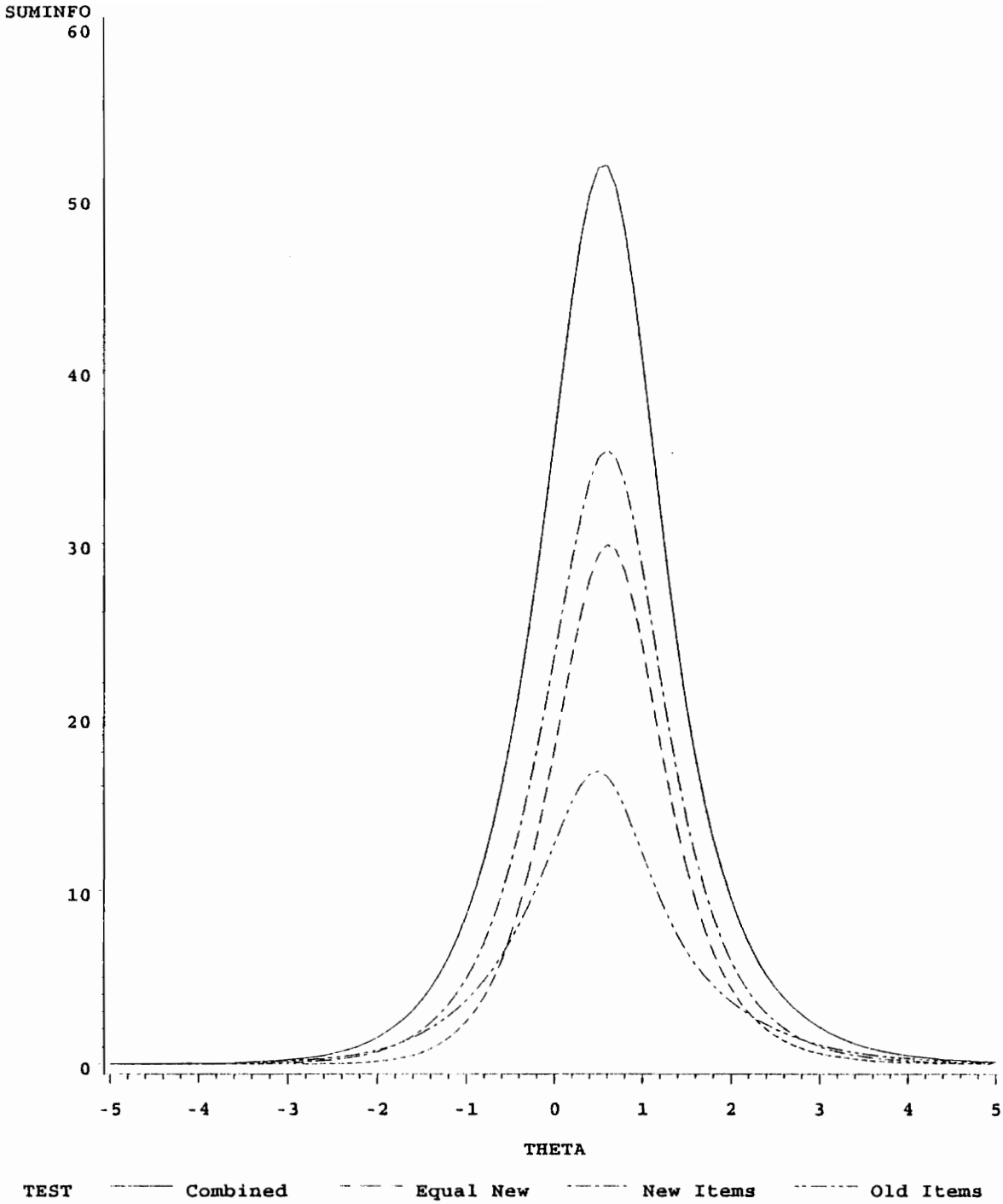

Figure 54: Test Information Functions for the S-N Scale. Test information functions for four different item pools within the S/N scale: (a) one item pool consisting of only the original items, (b) a second item pool consistingof all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

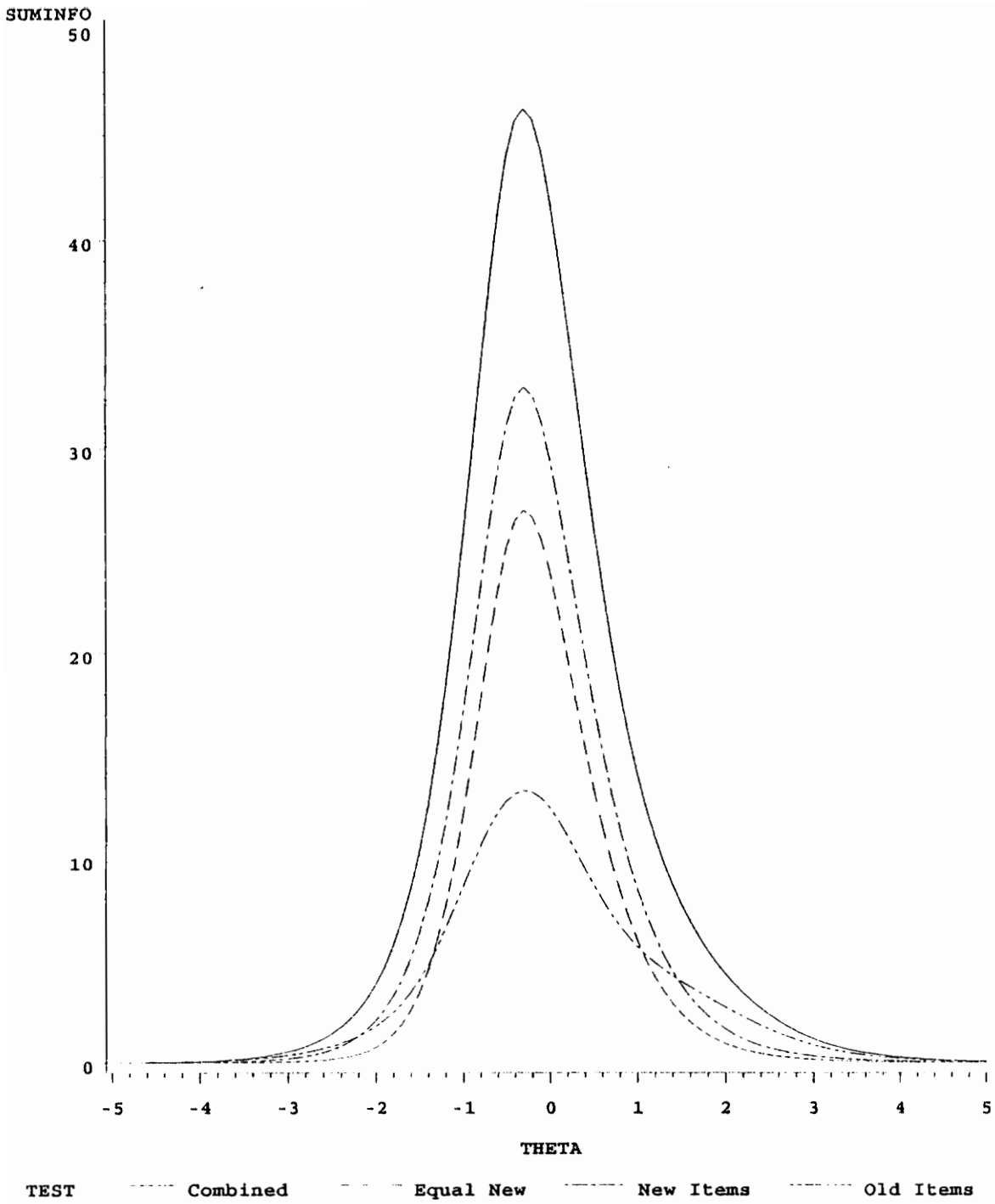

Figure 55: Test Information Functions for the T-F Scale. Test information functions for four different item pools within the T/F scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

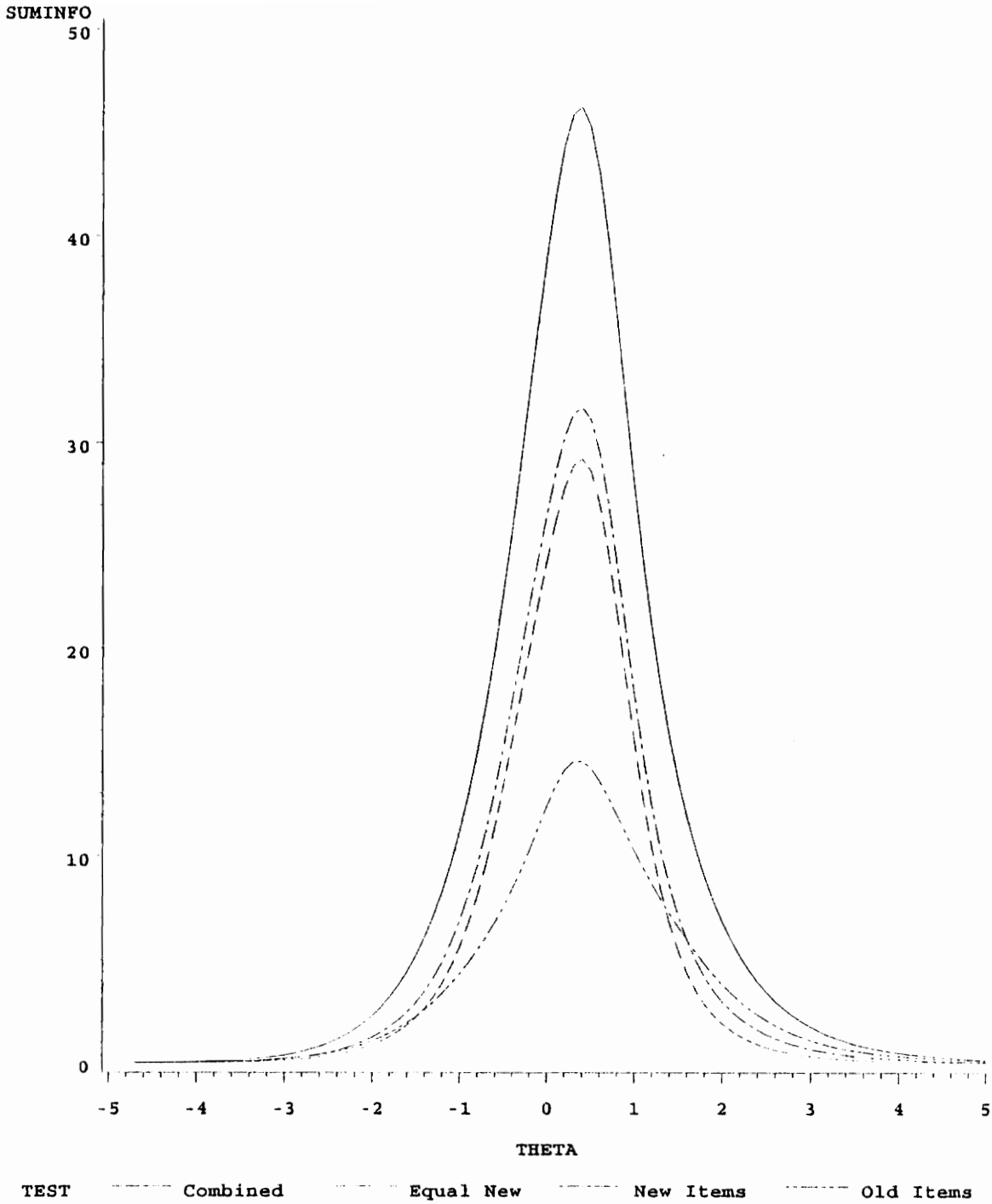

Figure 56: Test Information Functions for the J-P Scale. Test information functions for four different item pools within the J/P scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

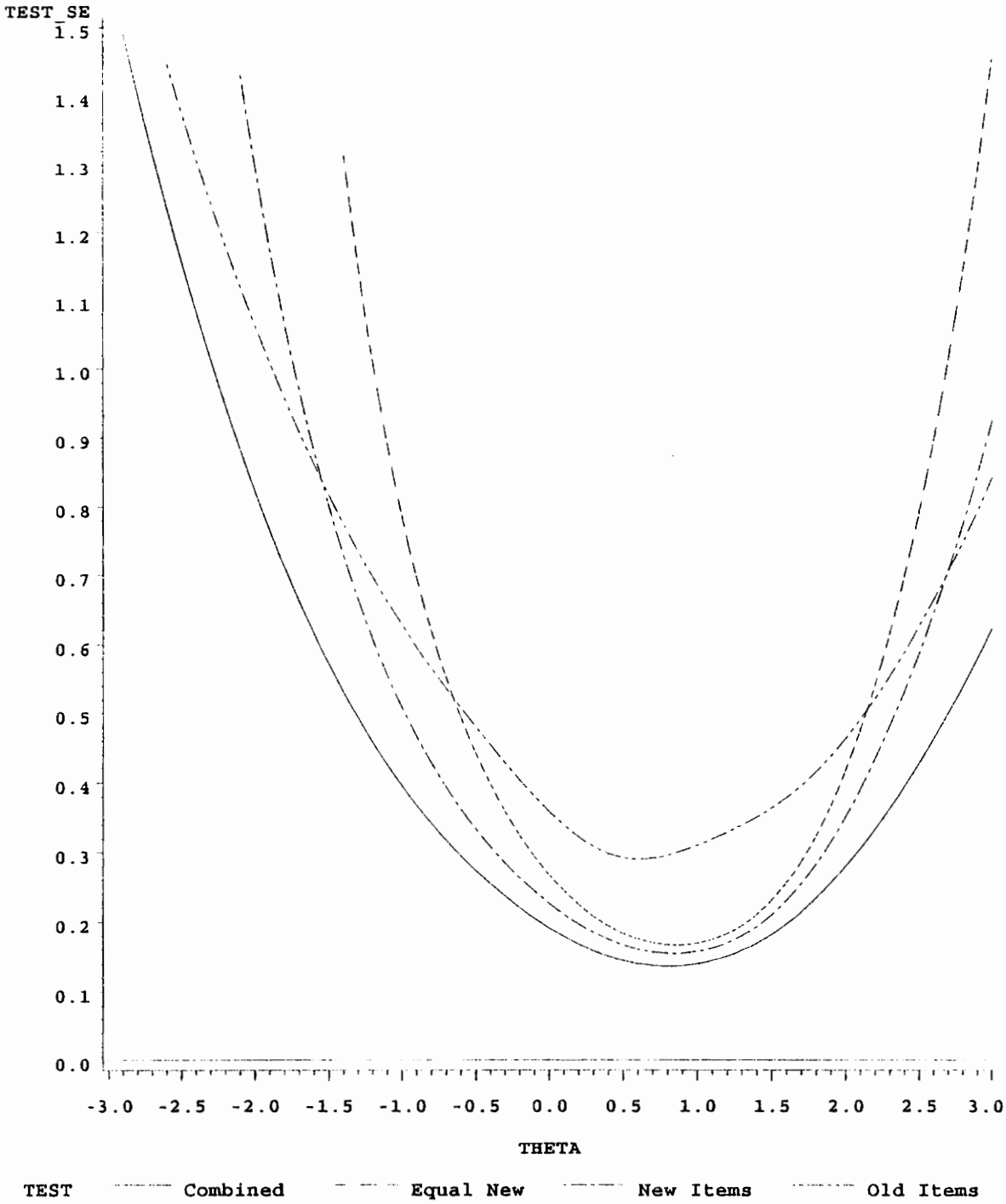

Figure 57: Standard Errors of Measurement for the E-I Scale. Standard errors of measurement (SEM) for four different item pools within the E/I scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

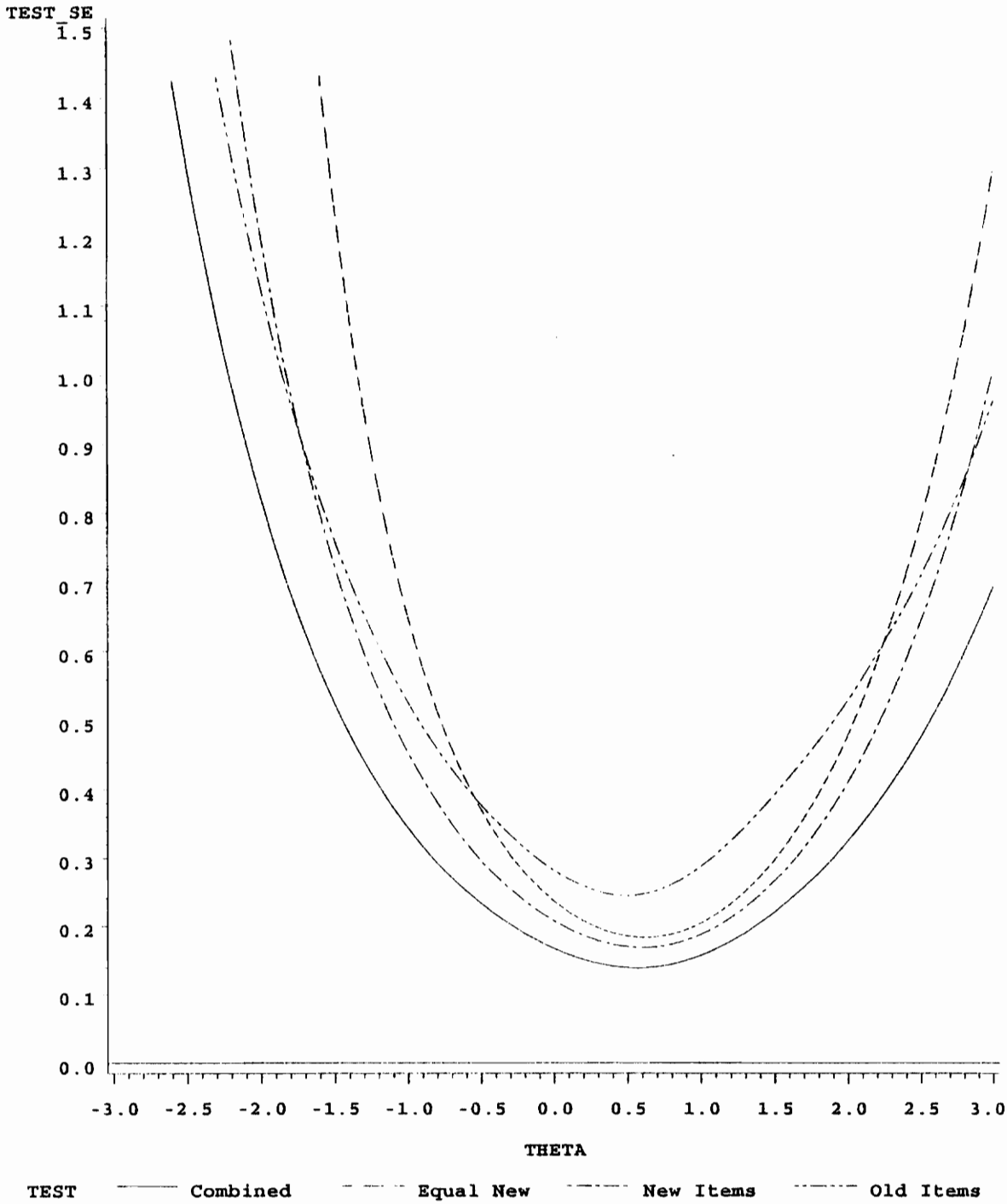

Figure 58: Standard Errors of Measurement for the S-N Scale. Standard errors of measurement (SEM) for four different item pools within the S/N scale: (a) one item pool consisting of only the original items, (b) a second item pool consistingof all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

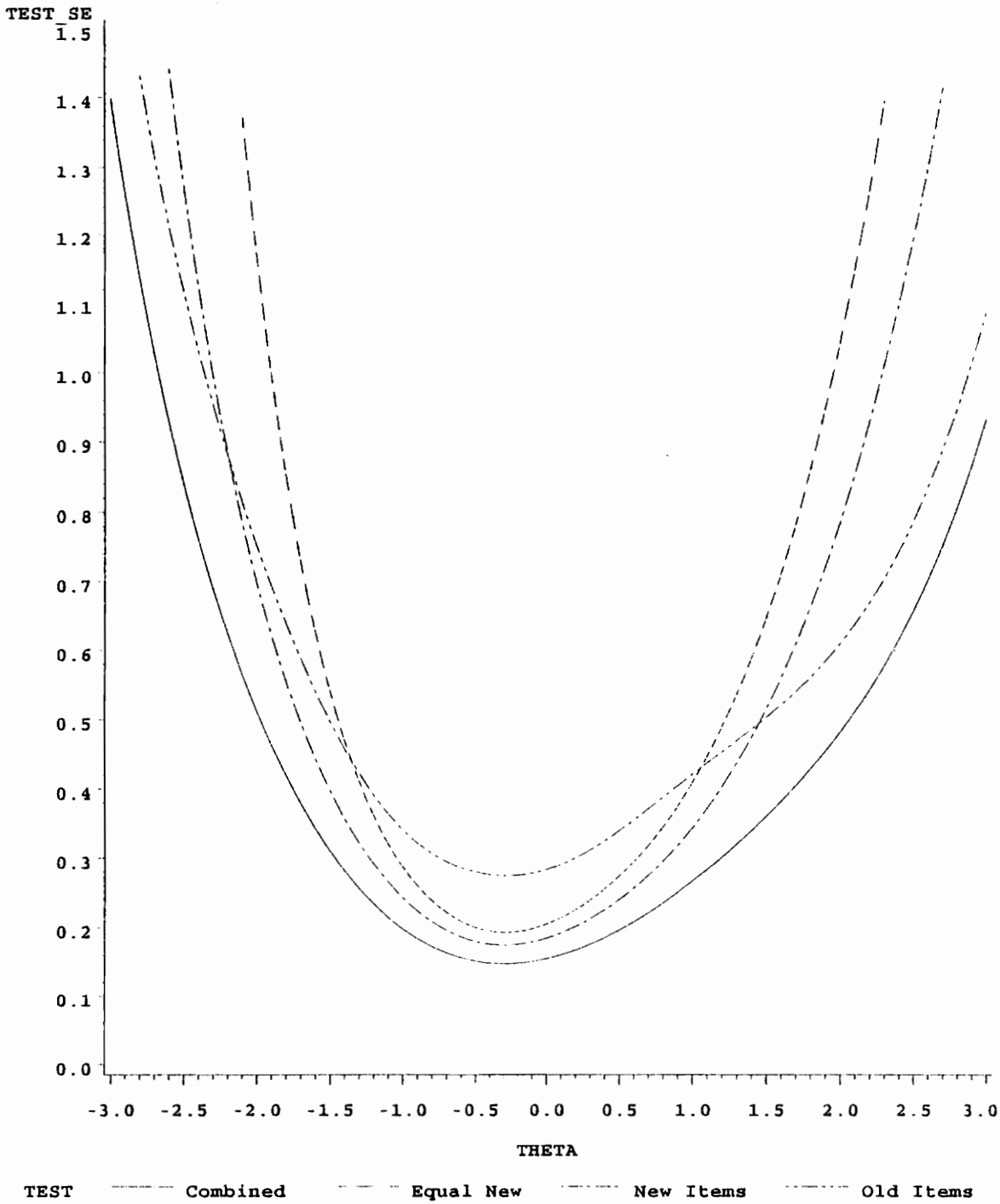

Figure 59: Standard Errors of Measurement for the T-F Scale. Standard errors of measurement (SEM) for four different item pools within the T/F scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

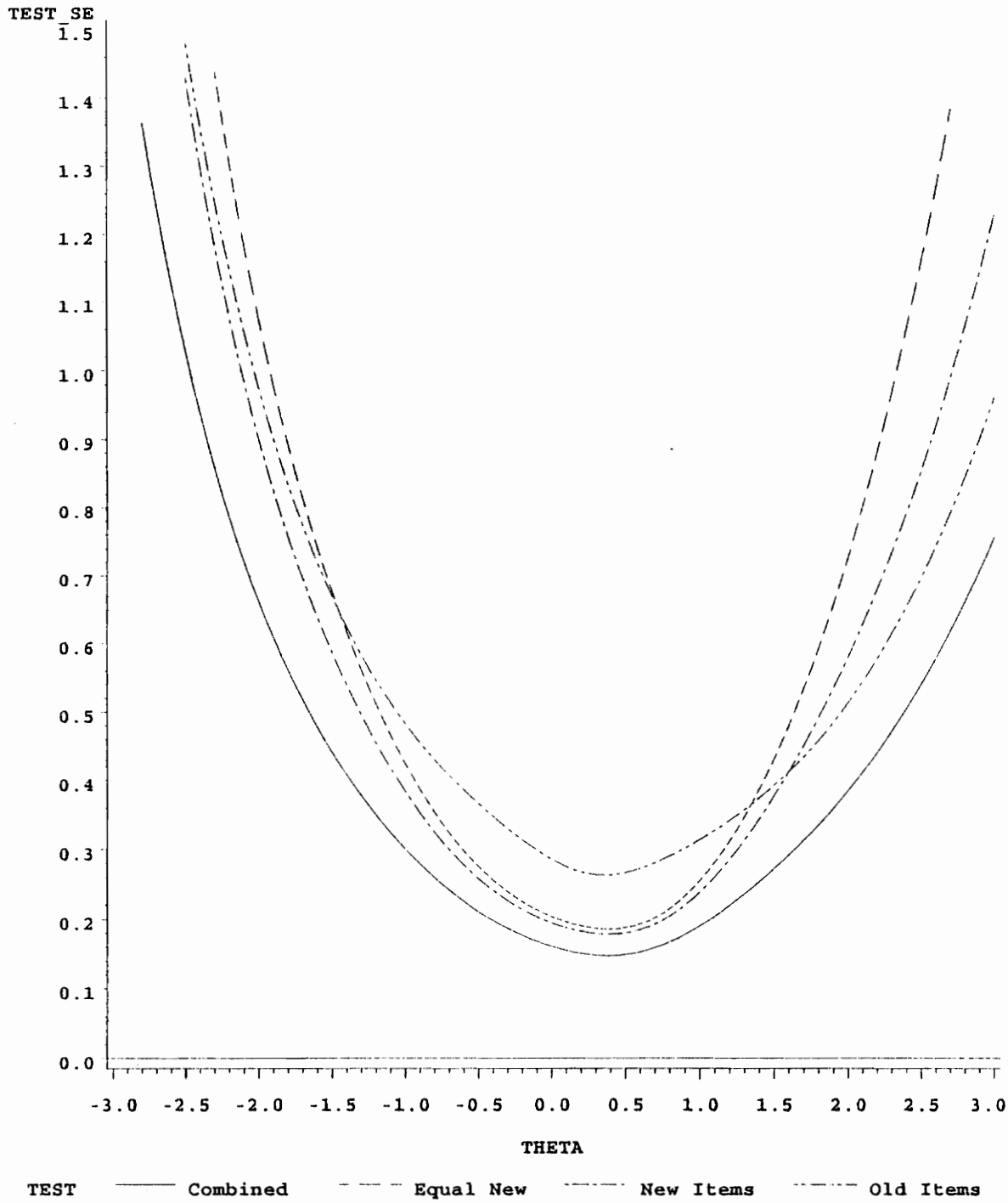

Figure 60: Standard Errors of Measurement for the J-P Scale. Standard errors of measurement (SEM) for four different item pools within the J/P scale: (a) one item pool consisting of only the original items, (b) a second item pool consisting of all of the new items, (c) a third item pool consisting of an equal number of new items as those contained in the original scale, and (d) a fourth item pool consisting of both the original items and all of the new items.

LESLIE ANNE THOMAS

LOCAL ADDRESS
1300 University Boulevard
Apartment 3111
Blacks burg, VA 24060
(703) 953-4052

PERMANENT ADDRESS
826 Oregon Avenue
Erie, PA 16505
(814) 838-6336

EDUCATION:

Master of Science Degree in Psychology, October 1994.
Concentration: Industrial/Organizational Psychology
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Thesis Title: Creating a Parallel Myers-Briggs Type Indicator Using Item Response Theory

Bachelor of Arts Degree in Psychology, May 1992.
Edinboro University of Pennsylvania, Edinboro, Pennsylvania
Graduated Summa Cum Laude

HONORS/AFFILIATIONS:

Presidential Honors Study Abroad Scholarship, Oxford, England, Summer of 1991
The Harrisburg Internship Semester Program, Fall of 1990, selected for highly competitive internship -- one candidate chosen from each state university
Presidential Honors Scholarship, 1989
University and National Dean's List, all semesters
Member of University Honors Program
Member of Alpha Chi National Honors Society
Member of Psi Chi National Psychology Honors Society

RESEARCH EXPERIENCE:

Virginia Polytechnic Institute and State University
Thesis Research - Robert J. Harvey, Ph.D. (Chair)
Created a test to parallel the substantive content of the Myers-Briggs Type Indicator (MBTI) to examine whether the new items could be written that measured the same underlying constructs as the original test and that substantially increased the amount of test information as compared to the original test in order to minimize type misclassifications and improve its test-retest reliability.

Virginia Polytechnic Institute and State University
Performance Appraisal Research - Neil M. Hauenstein, Ph.D. (Advisor)
Designed and implemented a field study to examine the influence of procedural justice factors - both formal (e.g., system) and informal (e.g., interpersonal) - on the employees' satisfaction with their evaluations, supervisors, and the appraisal system; Also examined the degree to which perceptual discrepancies, concerning the procedures utilized, between the supervisors and their subordinates added to the prediction of employee satisfaction variables beyond the employees' perceptions of the process.

Virginia Polytechnic Institute and State University
Research Assistant - Neil M. Hauenstein, Ph.D.
Assisted with assessment of Psychology Department as part of the University's Outcomes Assessment Program to determine how the undergraduate curriculum may be improved in order to better prepare its students and to compare the academic and career outcomes of psychology and non-psychology majors at Tech to those of students of other universities.

State Senate Republican Communications Office, Harrisburg, Pennsylvania
Staff Writer (Internship) - Wrote, helped implement, and oversaw a telephone survey to examine how negative political advertising influenced the public's recall and evaluation of political advertisements, as well as their evaluation of the sponsoring candidate and the candidate targeted in the advertisement.

PROFESSIONAL PAPERS:

Thomas, L. A. and Harvey, R. J., "Improving the Test Information Functions of the Myers-Briggs Type Indicator." Paper submitted to Society for Industrial-Organizational Psychology Conference, Orlando, Florida, May 19-21, 1995.

Thomas, L. A. and Harvey, R. J., "Improving the Measurement Precision of the Four Primary Scales of the Myers-Briggs Type Indicator." Manuscript submitted to the Journal of Personality Assessment, October, 1994.

PROFESSIONAL EXPERIENCE:

Personnel Services Department, Virginia Polytechnic Institute and State University
Research Assistantship, August 1993 - present
- Assisted with the position re-classification of 78 jobs within the University library system.
- Wrote and updated job descriptions for various departments throughout the University (e.g., Culinary Services, Personnel Services).

- Assisted with numerous workshops and presentations (e.g., leadership workshops, classification process workshops and presentations).
- Redesigned and implemented orientation program for all incoming faculty including conducting group orientation sessions, entering personnel information on various University systems (e.g., payroll, health insurance, etc.), and overseeing and implementing the Fall Benefits Sign-up for over 200 new faculty during August.
- Served on teams (e.g., Communications Team, Cobra Process Team, and New Employee Orientation Process Team) established to improve the efficiency of various personnel processes.
- Wrote procedures manuals for numerous programs (e.g., Savings Bonds Program, COBRA, Benefits Entry System (BES) for health insurance).
- Assisted with computer training and benefits training for Personnel employees.

RELEVANT COURSEWORK:

Research Methods
General Statistics
Organizational Psychology I and II
Advanced Psychometric Theory
Industrial Psychology I and II
Multivariate Methods of Analysis
Psychological Measurement
Cognitive Psychology
Multiple Regression Analysis

PROFESSIONAL INTERESTS:

Research interests include personnel assessment, selection, and placement, job analysis, performance appraisal, item response theory, psychometrics, and multivariate methods of analysis.

COMPUTER EXPERIENCE

Computer literacy with both PC and Macintosh computers and software packages including, but not limited to, Freelance Graphics, Harvard Graphics, Powerpoint, Wordperfect 5.1 & 6.0, Microsoft Word, Word for Windows, Amipro, ABC Flowcharter, Pagemaker, and SAS for Windows.