Order Number 9215046

The effect of contingency test instruction on locus-of-control and person reliability

Yang, Zijin, Ph.D.

University of Hawaii, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE EFFECT OF CONTINGENCY TEST INSTRUCTION ON LOCUS OF CONTROL AND PERSON RELIABILITY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF
THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN EDUCATIONAL PSYCHOLOGY

DECEMBER 1991

BY

ZIJIN YANG

Dissertation Committee:

Harold Ayabe, Chairman
Morris Lai
Mary Brandt
Sheldon Varney
Wimal Dissanayake
Michael Heim

ACKNOWLEDGMENTS

The writer wishes to thank his chairman, Dr. Harold Ayabe, for his constant encouragement and guidance in the preparation of this dissertation.

The writer also expresses his gratitude to the members of his committee for their helpful comments and advice: Dr. Morris Lai, Dr. Mary Brandt, Dr. Sheldon Varney, Dr. Wimal Dissanayake, and Mr. Michael Heim.

A special word of appreciation is extended to the educators and students in Changning District in Shanghai, China for their generous assistance in facilitating the writer's data collection in their schools.

Finally, my love and appreciation to my wife and daughter for their continual understanding and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Recently, the topic of person reliability, the consistency of a test taker, has begun to draw attention from people in the field of educational and psychological measurement. Several different methods have been proposed to produce person reliability indices (Sato, 1975; Harnish & Linn, 1981; Levine & Rubin, 1979; Ayabe & Heim, 1987). Researchers have applied person reliability indices to identify students with unique learning characteristics (Sato, 1975), to discriminate among students with different instructional backgrounds (Harnish & Linn, 1981), and to detect aberrant patterns of standard-setting judgments (Jaeger, 1988). More recently, Shishido (1991) applied person reliability to a college Japanese language course placement examination to ferret out students who might intentionally miss (sandbag) some of the test items.

Most of the studies on person reliability have focused on the introduction of new person reliability indices and their applications. A study associating person reliability with dispositional characteristics might be able to shed light on our understanding and application of person reliability. The purpose of this study is, first, to relate person reliability to locus of control, a selected dispositional characteristic, and second, to examine the effect of the test instructions on locus of control and person reliability.

The Development of Person Reliability

Scores obtained from a test are often used as the basis for decision making by researchers, administrators, and school teachers. An accurate interpretation of a person's test score in terms of reliability is a major concern in the field of educational and psychological measurement. To accomplish this, researchers have been working in two different but related areas: test reliability and the standard error of measurement (SEM) of a test and more recently, person reliability and the aberrant response pattern. Test reliability represents the consistency of the test instrument while the standard error of measurement is the standard deviation of the distribution of error scores which are the differences between subjects' obtained scores and their true scores. As test reliability increases, the standard error of measurement decreases. Person reliability represents the consistency of a test taker, or more specifically, the conformity of a person's response to the test items in relation to his/her group norm or a specific model. As the person reliability increases, one could expect this person's response pattern will be more similar to the group norm.

In classical test theory (Gulliksen, 1962; Thorndike, 1951; Lord & Novick, 1968; Nunnally, 1978), one basic assumption is that an observed test score consists of two parts: true score and error score. Parallel to this, the total variance has two components: systematic variance due to true score and random variance due to error. The ratio between the true (systematic) variance and the total variance (systematic and random variance) is an estimated reliability coefficient of the test for that group of test takers. The estimate of the reliability is traditionally used to calculate the standard error of measurement (SEM) for all test takers by multiplying the estimate of the standard deviation of test takers' observed scores by the square root of one minus the estimated reliability coefficient. The formula for calculating the standard error of measurement of a test is presented below:

$$SEM = s\,(1 - r)^{1/2}$$

where s = estimate of population standard deviation

r = estimated reliability coefficient

This standard error of measurement is used to calculate the confidence interval around one's estimated true score at a certain significance level.

The estimated standard error of measurement provided by classical test theory is descriptively accurate as an average for a group of test takers. However, it fails to reflect the variability of the measurement error at different score levels, e.g., low, average, and high, or some other specified score levels. Studies have shown that the standard error of measurement derived from classical test theory is not an adequate indicator in describing measurement accuracy at most score levels. Hence, the Standards for Educational and Psychological Testing written by the Committee of American Educational Research Association (AERA), American Psychological Association (APA), and National Council of Measurement in Education (NCME) to Develop Standards for Educational and Psychological Testing (1985) suggests that test publishers provide estimates of conditional standard error of measurements at critical score levels.

Several methods of approximating the standard error of measurement at specific score levels have been proposed, such as, the classical method (Thorndike, 1951), the polynomial method (Mollenkopf, 1949), the binomial method (Lord, 1955), the compound binomial (Lord, 1965), the variance components method (Hoyt, 1941), the IRT method (Lord, 1980), and Blixt and Shama's methods I-IV (1986). Lord (1955), for example, in proposing his binomial method, suggested the use of the observed score as an estimate of a person's true score, and he recommended correction for the known bias in the variance determined from finite samples when calculating the standard error of measurement. This approach leads to an estimate of the standard error of measurement at specific score levels for the total test. Lord's formula for the binomial method (1955) is presented below:

$$SEM = \{[X(K-X)]/(K-1)\}^{1/2}$$

where X = number of correct items

K = number of test items

Feld, Staffen and Gupta (1985) in their comparative study used Lord's binomial formula to calculate the conditional standard error of measurement at different score levels. The data set is comprised of over 1,600 ninth graders who took the listening part of the Iowa Tests of Educational Development (ITED). They calculated the standard error of measurement from individuals grouped into three-point total score intervals. They discovered that the conditional standard errors of measurement varied across score levels. Its distribution curve is a rainbow-shaped parabola, i.e., an arch viewed with the concave side facing downward. The largest values of the standard error of measurement (SEM) are located in the middle of the score continuum. The values of the standard error of measurement decrease gradually with the difference of the corresponding test scores from the group mean. The straight line, obtained from the single classical test theory based standard error of measurement for all persons and parallel to the horizontal axis, cuts through the "rainbow."

The distributions of the standard error of measurement obtained by Lord's binomial formula and the traditional formula based on the data provided by Feldt, Staffen, and Gupta (1985) are presented in Figure 1, which shows that the maximum value of the standard error of measurement is more than twice the minimum. The largest standard error of measurement at score level 22-24 is 3.26, and the two smallest standard error of measurements at score levels 1-3 and 43-45 are 1.45 and 1.51 respectively. The traditional standard error for the whole test is 2.81. This phenomenon implies that the standard error of measurement computed by the traditional formula for the test as a whole does not

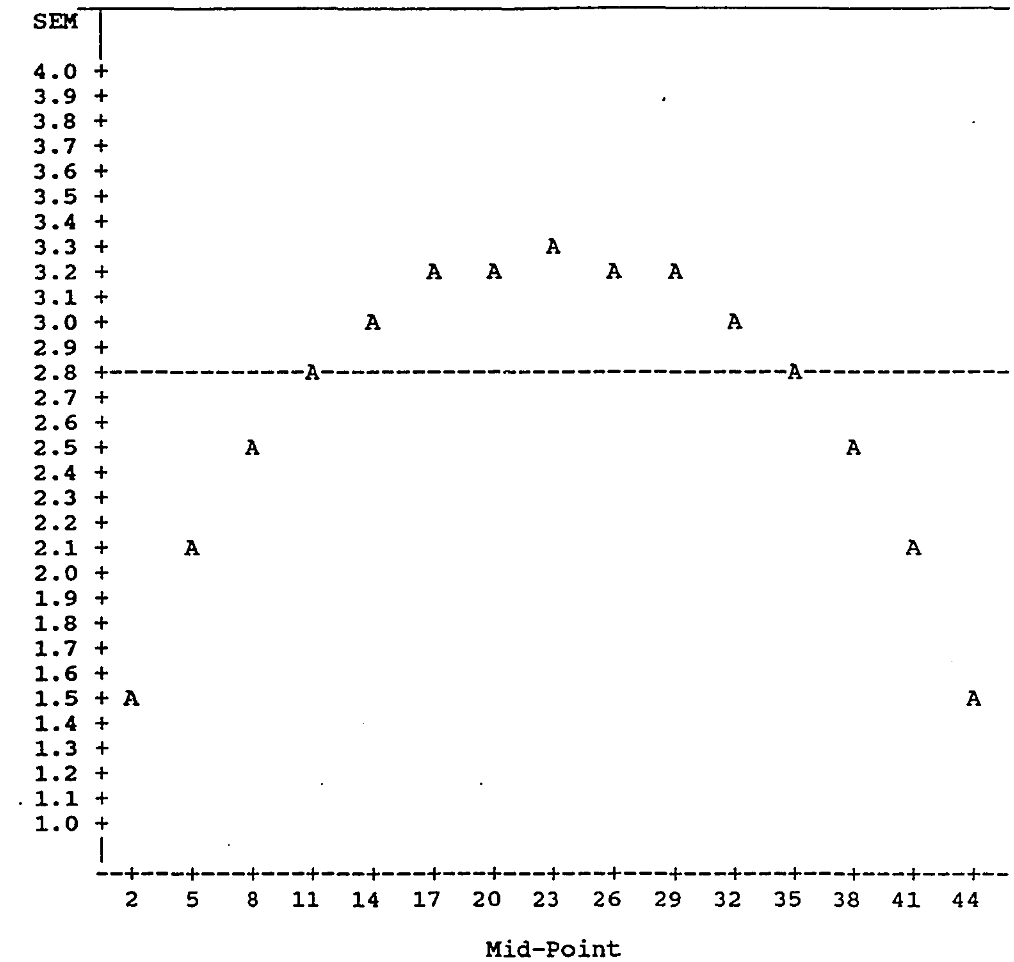


Figure 1
Comparison of the Traditional Standard Error of Measurement
and the Conditional Standard Error of Measurement
at Specific Score Levels

* From Feldt, Staffen and Gupta (1985) based on the data collected from 1,600 ninth graders on the listening part of the Iowa Test of Educational Development (ITED)

adequately represent the errorfulness of many test takers, especially those with scores in the middle or at the two extremes of the score continuum. It would be more appropriate to use the conditional or specific standard error of measurement applicable to the test taker's score level, instead of the overall standard error of measurement to calculate the confidence interval, when including a confidence interval along with a person's test result in a report.

It is now reasonable to make a statement that the conditional standard error of measurement at a specific score level offers more precise information of test takers' ability than the traditional standard error of measurement for the whole test. This conclusion will naturally lead one to take in this direction one step further by asking the following questions: If so, does every test taker at the same score level, assuming there is more than one person at that level, respond to the test items the same way? In other words, is a conditional standard error of measurement at a specific score level "good enough" for every test taker at that level? If not, is it possible to derive a better indicator which will provide unique information about a person's response to the test items so that a more accurate interpretation of a person's test performance and score can be made with reasonable confidence?

It is observed that test takers tend to answer test items differently. The way that a given individual responds to the test items would most likely vary, sometimes substantially from the response pattern of another individual with the same test score. From a theoretical point of view, a 10-item test can yield 1024 different response patterns. Even with a test score of three points (i. e., three items answered correctly) on a 10 point test, there could be as many as 120 response patterns. Although some extreme response patterns may never occur in real life, there would still be large variability among the response patterns that do happen. In other words, some persons would miss the easy items and/or answer the difficult items, thus producing an aberrant response pattern. The information included in each individual's response pattern, if utilized, can be of great assistance to researchers and classroom teachers who can not make sense of the right and wrong answers for various reasons in determining and interpreting the meaning of a test score.

Measures of Person Reliability

A number of measures of person reliability have been proposed by researchers. These measures can be classified into three groups according to the theory or model on which they are based. They are item response theory-based appropriateness indices, group dependent response indices, and classical test theory based index.

The first group of measures is based on the item response theory (IRT) model which is purported to be "sample free" or "group independent" (Hambleton & Swaminathan, 1985). The IRT model based person reliability indices measure the goodness of fit between a person's response pattern and a specific psychometric model derived from IRT theory. The IRT model based person reliability indices or appropriateness indices include the Maximum Log Likelihood Function (lo), (Levine & Rubin, 1979), the Standardized Maximum Likelihood Function (z_3) (Levine & Drasgow, 1982), the Squared Standardized Residual (Wright, 1977; Wright & Stone, 1979), Likelihood Function Curvature Statistics (Jackknife variance estimate) (Mosteller & Tukey, 1968), Item-Option Variance (IOV) and Fit Statistics (F1, F2) (Runder, 1983), Unweighed Fit Statistics, U1 & U2, (Wright & Panshapakesan, 1969) and the (Standardized) Extended Caution Index (ECI, ECIZ) (Tatsuoka & Linn, 1983). In a series of comparative studies, researchers have discovered that lo is the simplest index and has great capacity for detecting aberrant response patterns in most of the cases and, thus, is recommended for use if an appropriateness index is necessary for decision-making.

The second group of measures is based on the test takers' pattern of right and wrong answers. The indices derived from these methods indicate how atypical a person's response pattern is, relative to the typical response pattern of the whole group. These indices include the Personal Biserial Correlation (r) (Donlon & Fischer, 1968), Caution Index (CI) (Sato, 1975), the Modified Caution Index (MCI) (Harnish & Linn, 1981), the Agreement (A) and Disagreement (D) Indices (Kane & Brennan, 1980), the Dependability Index (Θ) (Kane & Brennan, 1980), the van der Flier index (U) (van der Flier, 1982), and the Norm-Conformity Index (NCI) (Tatsuoka & Tatsuoka, 1982).

In their comparative study, Harnish and Linn (1981) used the above ten response pattern indices to analyze data collected from high schools in the 1978 annual statewide survey, known as the Illinois Inventory of Educational Progress (IIEP). The results indicate that these ten indices all have high intercorrelations except for one, the agreement index, and low correlations with test takers' total score. Among them, the MCI is the one which has the lowest correlation (-.21) with test takers' test scores. It indicates that MCI is least redundant with total test scores, i. e., it offers information not included in the total score. This property of MCI is desirable and therefore of interest.

The third kind of person reliability index of approximating the consistency of a person's response pattern is the Person Average Reliability (PAR). It is proposed by Ayabe and Heim (1987) based on the classical test theory. This index indicates how similar a person's response pattern is to the response patterns of the rest of the test takers.

The connection between IRT model based measures and group dependent measures is discussed by Tatsuoka and Linn (1983). They demonstrated the existence of some correspondence between Sato's Student-Problem curve theory and group response curves of item response theory. The connection between the group dependent response pattern measures and the classic test theory based response pattern measure also have been noted by Ayabe and Heim (1987).

In the following section, lo, MCI, and PAR, representatives of the three above-mentioned approaches of measuring person reliability or response pattern appropriateness, will be discussed in detail.

The Maximum Log Likelihood Function(lo)

Appropriateness index lo is a measure of aberrant response pattern based on item response theory.

The basic assumption of item response theory is that in a test situation a test-taker's performance can be predicted item by item by defining a person's characteristic referred to as latent trait (Θ). The relationship between observed performance and unobservable traits is a stable mathematical function. According to this theory, the estimate of a test taker's ability is purported to be "independent of the particular sample of test items that are administered to the test taker." Similarly, the description of a test item is purported to be "independent of the particular sample of test takers drawn from the population." A statistic indicating the precision of each test taker's ability estimate is provided, and it is "free to vary from one test taker to another" (Hambleton & Swaminathan, 1985).

Item response theory was first formally introduced by Lord in 1953, when he developed an item response model and related methods for parameter estimation and applied a normal ogive model to real achievement and aptitude test data. Later, extensions of item response theory for analysis of different data sets (dichotomous vs. polychotomous) and dimensional models (unidimensional vs. multidimensional) were proposed (Samejima, 1981).

In the 1970's and 1980's, thanks to the availability of high speed computers, important breakthroughs were achieved in many areas (Lord, 1980; Wright & Stone, 1979; Levine & Rubin, 1979; Weiss, 1973). Some of the promising applications include test score equating, test design, test evaluation, and, more recently, appropriateness measurement.

The goal of appropriateness measurement is to identify test takers with inappropriate response patterns, i.e., with a relatively large proportion of difficult items answered correct and easy items wrong. It takes two steps to achieve this goal. First, a general psychometric model is fitted to a large sample of average test takers. Then, an appropriateness index is used to measure the degree to which each person's response pattern fits the model used to characterize normal behavior.

The appropriateness measure lo, the Maximum Log Likelihood Function, is proposed by Levine and others (Levine & Rubin, 1979; Drasgow, 1982). It is based on the notion that an appropriate response pattern should be representative of the group whose abilities are measured by the test. Mathematically, lo is the natural logarithm of the three-parameter logistic likelihood function evaluated at $\hat{\theta}$, the maximum likelihood estimated (θ). To derive lo, the maximum likelihood procedures are used to estimate item parameters and test takers' abilities. Thus, lo is the measure of a test taker's contribution to the likelihood function. In this sense, an aberrant response pattern is one which contributes little to the maximum likelihood function. Suppose that a person has an aberrant response pattern yielding an answer sheet with incorrect responses to a number of easy items and correct responses to several difficult items. The likelihood of this response pattern will be very small for all values of $\hat{\theta}$, and consequently the maximum of the likelihood function will be very small. For this reason a small value of lo suggests an aberrant response pattern.

In the first large scale, systematic appropriateness measurement study, Levine and Rubin (1979) created answer sheets with spuriously high and spuriously low scores by modifying simulated item response data. The created data set was used to calculate the Maximum Log Likelihood Function lo and some other item response theory based appropriateness indices. The results showed that lo, a response pattern appropriateness index, can classify normal and moderately aberrant response patterns fairly well.

Levine and Drasgow (1982) extended the above initial results to more diversified and realistic conditions. This time they used actual and simulated data to study the effects of the inclusion of aberrant test takers in samples of average test takers. They also found that lo and other item response theory based appropriateness indices can identify aberrant response patterns with both actual and simulated data.

The formula for computing the Maximum Log Likelihood Function lo is presented below:

$$lo = \sum_{i=1}^{n} [U_i \log P_i(\hat{\theta}) + (1-U_i) \log Q_i(\hat{\theta})]$$

where U_i is the dichotomously scored (1 = correct, 0 = incorrect) item response for item i (i=1, 2,..., n), and

$P_i(\hat{\theta})$ = probability of a correct response to item i among examinees with ability θ.

$$P_i(\hat{\theta}) = c_i+(1-c_i)/\{1+\exp[-Da_i(\theta)-b_i]\}$$

where

a_i is the item discriminating power,

b_i is the item difficulty,

c_i is the lower asymptote of the item characteristic curve (ICC), and

D is a scaling constant usually set equal to 1.702

$Q_i(\hat{\theta}) = 1-P_i(\hat{\theta})$

$\hat{\theta}$ = ability estimate

The Modified Caution Index (MCI)

MCI is an index of person reliability, or rather person unreliability, characterizing how different a person's response pattern is in relation to that of the whole test group. It is a modification of Sato's caution index (CI).

Sato's theory is built on Guttman's model (1941) that there exists a latent or underlying continuum to which the response to a particular item is to be related. Sato also believes that it is possible to yield a Guttman scale, i.e., order the items such that, "persons who answer a given question favorably all have higher ranks than persons who answer the same questions unfavorably." The perfect Guttman scale is not likely to be found in practice, but satisfactory approximations to it are still possible. Based on this assumption Sato (1975) developed a student-problem (S-P) chart. Figure 2 is an example of this chart. It is a matrix in which rows are ordered by descending total number of correct responses; columns are ordered by ascending order of difficulty level of items. The student-curve is an ogive reflecting test-takers' performance represented by their total scores; the problem-curve is, on the other hand, an ogive reflecting item difficulties. Sato argues when a person's responses to the test items form a Guttman scale, i.e., answer all the easy items correctly and miss only the

Rank	Item 1	2	3	4	5	Correct Total	Caution Index	Modified Caution Index
1	1	1	1	1	0	4	0.00	0.00
2	1	1	1	0	1	4	0.65	0.33
3	1	1	1	0	0	3	0.00	0.00
4	1	1	0	1	0	3	0.16	0.08
5	1	1	0	0	1	3	0.65	0.31
6	1	0	1	0	1	3	1.13	0.54
7	1	1	0	0	0	2	0.00	0.00
8	1	1	0	0	0	2	0.00	0.00
9	1	0	1	0	0	2	0.44	0.23
10	1	0	0	1	0	2	0.59	0.31
11	0	1	1	0	0	2	0.74	0.38
12	0	1	0	1	0	2	0.88	0.46
13	1	0	0	0	0	1	0.00	0.00
14	1	0	0	0	0	1	0.00	0.00
15	0	1	0	0	0	1	0.45	0.22
16	0	0	1	0	0	1	1.14	0.56
17	0	0	0	1	0	1	1.36	0.67
18	0	0	0	1	0	1	1.36	0.67
$n_{.j}$	12	10	7	6	3			

Figure 2

Sato's Student-Problem Chart for Five Items and Eighteen Students

* From Harnish and Linn (1981) based on hypothetical data

difficult ones, then the student and problem curves will grow increasingly different.

Sato developed the caution index (CI) according to the student and problem curves. In order to avoid large values of CI for those at the two extremes of the total score continuum, Harnish and Linn (1981) introduced the Modified Caution Index (MCI), a standardized version of CI, with the upper limit of 1 and the lower limit of 0. A small MCI indicates that a person's response pattern is similar to that of the group norm, while a large MCI implies that this person's response pattern is quite different from the group norm. Harnish and Linn (1981) suggest that an MCI of 0.3 should be used as the cutoff point in distinguishing normal and aberrant response patterns, if necessary. The formula for calculating MCI is presented below:

$$MCI = \frac{\sum_{j=1}^{n_i} (1-U_{ij} n_{.j}) - \sum_{j=J+1 \ldots n_j}^{J} U_{ij} n_{.j}}{\sum_{j=1}^{n_i} n_{.j} - \sum_{j=J+1 \ldots n_j}^{J} n_{.j}}$$

where i = 1, 2, 3, ... I, indexes the person

j = 1, 2, 3, ... J, indexes the item

$u_{.j}$ = 1 if examinee i answers item j correctly,
0 if examinee i answers item j incorrectly

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

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

CI and MCI have been used in a variety of studies both in Japan (Sato, 1975) and in the United States (Harnish, 1983). Sato suggests that the response pattern index may be used along with the total test score to identify students of different types. Table 1 shows the classification of students into four groups: A, who are academically doing well, B, who make careless mistakes, C, who need to study more, and D, who are doing poorly at school in terms of academic achievement and study habit. With this information at hand, researchers and classroom teachers will be able to refine their study or take proper measures to guide the students.

MCI has been applied in a statewide survey in Illinois (Harnish & Linn, 1981). The results of a hierarchical ANOVA with school as the unit of analysis indicate that region is the first factor and school is the second factor in determining the value of students' MCI. The significant school-within-region effects indicates that the response patterns of the students in school within certain areas of the state are different from those of students in other

Table 1

Classifying Students According to the Modified Caution Index (MCI) and Test Score

		MCI	
		low (less than or equal to .3)	high (greater than .3)
Test Score	high (greater than 50% of items correct)	A	B
	low (less than 50% of items correct)	C	D

areas. They speculated that curriculum offerings might contribute to these large differences.

Similarly, MCI has also been applied in identifying classes with atypical instructional background that change the relative difficulty ordering of the test items. More commonly, MCI has been used to identify students for whom the standard interpretation of a test score may be biased or misleading (Tatsuoka & Tatsuoka, 1983). Students at a same score might differ in their response pattern. These procedures also have been applied to the problem of screening judges in standard-setting applications (Jaeger, 1988).

The Person Average R (PAR)

Person Average R (PAR), a more recent measure of person reliability or test response pattern is proposed by Ayabe and Heim (1987) based on classical test theory. It describes how similar the way a person responded to the test items in relation to the other test takers.

PAR is obtained by calculating the average correlation of a person's response pattern with that of every other test taker. It is a variation of Kuder-Richardson formula 20 (KR_{20}). In essence, an estimate of internal consistency reliability of a test derived from KR_{20} is the mean of all possible split-half correlation coefficients, i.e., the mean of all item reliabilities. Kuder and Richardson (1937) suggested that researchers" use the average correlation of item *i* with the *n*-1 other items of the test as an estimate of the reliability of item *i*." The computation of PAR for an individual test taker is the same as computing the item reliability coefficient except the data, using an item-by-person response matrix instead of a person-by-item matrix.

To obtain the reliability of a specific item, one starts with a person-by-item matrix. First, The test takers' responses to that item are correlated with their responses to the rest of the items. Then, correlation coefficients are summed up and divided by the number of correlations.

Analogous to this, to obtain the person average R, the only change that needs to be made is to transpose the person-by-item matrix into an item-by-person matrix, correlate a person's response to the test items with that of the other test takers, and then obtain this person's PAR, the mean of all possible correlation coefficients by averaging the sum. If a person responds to the test items in a way similar to others, one can expect many high correlation coefficients and consequently a high mean or a high PAR.

PAR in theory can range from -1 to +1. Generally positive correlations are expected though small negative correlations are still possible. Since PAR represents the degree of consistency of a person's response pattern with that of the rest of the test takers, a large PAR indicates that this person responded to the test in a fashion similar to the rest of test takers. In other words, a person picks up the items that most of the test takers answer correctly and only misses those that his/her peers also fail to get correct.

A pilot study by Shishido (1991) has indicated that PAR has high correlations with most of the major response pattern indices, e.g., -0.99 with MCI, and a low correlation (0.25) with total score. Compared with lo and MCI, PAR is conceptually straightforward and relatively easy to calculate with the use of a computer and a simple correlation procedure without requiring additional computer software or undergoing complicated programming as the computation of MCI and lo.

The formula for calculating person average R is presented below:

$$PAR_i = \sum_{j=1}^{N-1} r_{ij} / (N-1)$$

where $i \neq j$

N = number of items

r_{ij} = correlation coefficient between two items

Internal-External Locus of Control

It has been shown in the previous section that a person with indices indicating low person reliability or an aberrant response pattern tends to react to the test items in ways contrary to the main body of test takers, i. e., miss easy items and answer difficult items. In this section, a dispositional characteristic, locus of control, is selected among other contributing factors as the focus of the present study. It is proposed here that a person's locus of control orientation, whether one believes that his/her chance or effort will determine the outcome of an event, might relate to his/her test-taking behavior represented by person reliability indices.

Definition of Locus of Control

The theory of locus of control was developed in the 1950s. The largest body of empirical data about 'perceived control' is derived from Rotter's social learning theory (1954). In his 1966 monograph, Rotter provides the definition, theoretical foundation, and general characteristic of the concept of locus of control. Rotter defines it as a "generalized expectancy primarily associated with past experiences." Locus of control describes a person's perception of the contingency between his/her behavior and the reinforcements he/she receives.

Rotter observes that people have developed generalized expectancies in learning situations with regard to whether or not reinforcement, reward, or success in these situations is dependent on their own behavior or is controlled by external forces, e. g., luck, chance, or experimenter control. If reinforcements are seen as outcome of efforts, locus of control is considered "internal." If, however, they are primarily attributed to events or forces outside the person's control (e.g., chance, luck, and fate), locus of control is considered "external." The difference between internal and external control does not lie in the source of reinforcement but in the *perceived* source of reinforcement. Rotter points out that such expectancies regarding control generalize across a large variety of situations. Rotter suggests that differences in learning and other behaviors could be interpreted according to the subjects' locus of control orientation.

Locus of Control and Person Reliability

The construct of locus of control gained great popularity in the 70's, and extensive reviews have been conducted by Rotter (1966) and Lefcourt (1966, 1976, 1984). The construct of locus of control has been employed in studies of individual and group differences in many areas among which are academic achievement, mood, risk-taking behavior, impulsivity, decision-making approaches, perception, and information processing. The results from these studies are inconclusive and sometimes mixed. The general tendency is that external locus of control orientation appears to be related to abnormal, irregular, unusual, and even undesired behaviors or performances, such as underachievement, impulsivity, inattention, short retention of information, and others.

If an individuals' locus of control influences his/her performance in a variety of learning situations, then it is reasonable to expect that an individual's locus of control will affect his/her test-taking performance. In other words, if an individual tends to believe more strongly that chance determines the test result (external orientation), he/she would probably have a less predictable sequence of right and wrong answers to the test items, namely an aberrant response pattern, while those with a stronger belief that one's own effort determines the outcome of an event (internal orientation) would most likely respond to the test problems in a more consistent and predictable way.

It can also be argued on the basis of the following parallel findings that the aberrant response pattern is related to the external control orientation. Among the behaviors which have been associated with locus of control are the following two behaviors: risk-taking and decision-making behaviors. Both excessive risk-taking and hasty decision-making behaviors resemble to the characteristics of aberrant response pattern, which support the proposition that locus of control and person reliability are related.

First, the findings from studies relating risk-taking behavior to locus of control provided evidence directly relevant to the assumed relationship between locus of control and person reliability. Liverant and Scodel (1960) conducted a study to examine the risk-taking behaviors of the internal and external subjects. Eighty-five students of an introductory psychology class at Ohio State University participated in the experiment. They were given some money for betting, instructions on how to play the dice game, and an explanation of the payoff matrix detailing the objective probabilities and payoff for each of the bets. The bets can be generally classified into three categories: high, intermediate, and low probability.

The results of the Mann-Whitney test indicate that internal subjects chose statistically significantly more intermediate probability bets than external subjects (z=2.89, $p<0.002$) and correspondingly internal subjects chose fewer low probability bets than external ones (z=1.68, $p<0.05$). Internal subjects wagered a greater amount of money on their bets than external subjects. Specifically, internal subjects wagered more money on their safe (intermediate probability) bets than on the risky (low probability) ones. The external subjects wagered more money on the high probability bets which only offered a return less than the amount they had bet. The external subjects also bet more frequently on low probability choices. To say it differently, internal subjects prefer safe choices and bet money consistent with their choices, confident in winning at least some of them. External subjects make contradictory decisions, making risky choices, but without much confidence in their decisions, and they bet big money on choices of high probability with only a small return. Liverant and Scodel's finding was confirmed by Julian, Lichtman, and Ryckman's (1968) study of dart-throwing games. They reported essentially the same finding that internal subjects preferred choices with high probabilities of success while external subjects preferred choices with relatively low probabilities of success.

The above findings suggest that external subjects tended to bet unpredictably. Their greater variability in making choices and their irregular betting practice lead to the expectation that external subjects make their choices in a more random manner. On the other hand, as a function of their belief in their efforts, internal subjects tended to make better use of betting strategies. Internal subjects were more restricted in their choice of bets, selecting the choices with a better chance of being correct on an objective basis.

When playing dice games described above, internal subjects tend to choose, with confidence, "safe" bets with high probability of being correct according to the payoff matrix, and external subjects choose bets, safe or risky, at random. Likewise, in a test-taking situation, internal subjects when faced with difficult multiple-choice questions would act in a similar way, using better test-taking strategies, concentrating on the information provided, eliminating improbable choices, and systematically choosing more safe answers with a relatively higher probability of being correct. External subjects might approach the same test in a different way, randomly choosing answers with a wider range of probability of being correct, including improbable choices. It follows that since internal subjects have more safe choices their response patterns would be more consistent with regard to the difficulty levels of the test items determined by the whole test group, and external subjects would have deviant or aberrant response patterns.

Another area of study which lends support to the notion that internal and external control is related to aberrancy is decision-making studies. The first study by Rotter and Mulry (1965) asked the subjects to do a difficult matching task under two conditions: skill and chance instructions. Results indicated that internal subjects spent more time deliberating about their decisions than did external subjects especially when receiving skill instruction. External subjects were more deliberate with chance directions but the difference was not significant.

The above study was replicated by other researchers with some modifications and obtained essentially the same results reflecting the different approaches by internal and external subjects. Julian and Katz (1968) compared the variability of internal and external subjects' decision-making time with problems of differing difficulty levels. They reported that internal subjects took increasingly more time to make decisions as the difficulty level of the problems increased. External subjects did not vary as much with the difficulty level of the problems, behaving as if there were no differences between simple and difficult choices. To say it in another way, since internal subjects believe that their effort can make a difference, they become more deliberate and work harder in their attempt to solve problems, especially if a task challenges their competence. On the other hand, external subjects with a lack of belief that personal endeavor could make a difference do not seem to demonstrate clear differences about different tasks as internals do.

Based on the above findings, it can be reasoned that because of the different decision-making approaches, internal test takers' response patterns will differ from those of external test takers. Internal subjects may tend to take more time in test-taking, concentrate on problem solving, be very careful in test-taking, examine the probability of being correct for each choice, and exercise caution in making decisions. On the other hand, external subjects tend to treat all problems the same way, making quick decisions at random regardless of their difficulty levels. Consequently, they choose some correct and some incorrect answers in an irregular fashion not consistent with the order of items' difficulty levels, and finally, produce aberrant response patterns.

There are other reasons to support the assertion that locus of control and response pattern are associated. Ryckman and Sherman (1973), and Kilpatrick, Dubin, and Marcotte (1974) all found locus of control correlates with measures of moods, such as "feeling of inadequacy," and "vigor." Kilpatrick, Dubin, and Marcotte (1974) asked students enrolled in each of four years of medical school to report their moods. Internal medical students reported less mood disturbances on the Profile of Mood States (POMS) than their more external students. Internal subjects described themselves as "less tense, less anxious, less depressed, less fatigued, and less confused" than did external subjects in each of the four classes (first year through fourth year). Moreover, while no differences with regard to 'vigor' were obtained in the first year and fourth year classes, external subjects reported feeling markedly less vigor than did internals in the second and third year classes. According to their study, the second and third year of medical school are generally considered to be the most stressful and uncertain of the four years in medical school.

Internal students remained buoyant, vigorous, and energetic under continuously uncertain and more stressful circumstances, in contrast to external ones who reported a marked decline in terms of vigor and energy. Tension, anxiety and depression can be interpreted as debilitating and undermining and vigor as facilitating and stimulating. The above reasoning indicates that internal subjects are likely to remain active in problem solving and external subjects are more likely to yield to demanding situations and give up when confronted with challenges. What is being suggested here is that locus of control is pertinent to the strategies of coping with challenges. An internal subject will persist in the belief that he can deal with difficult situations well, and will strive after valued goals in the belief that his efforts are meaningful and effective. As for the external subjects with a dominant negative past experience, they tend to be less active and probably more impulsive in response to frustration when confronted with challenge. Test-taking usually is a very demanding situation, especially when dealing with difficult problems. Essentially, how a test taker performs under this circumstance will largely determine his/her response pattern. Following this line of reasoning, it appears safe to hypothesize that a person in a steady, vigorous mood will behave in a more systematic way when compared with a person under constant tension. Here is again evidence supporting the postulation of the relationship between locus of control and person reliability, or rather the relationship between low person reliability or aberrant response pattern and external locus of control orientation.

In sum, according to Rotter's definition of locus of control, people can be categorized into two groups: internal and external subjects. As a function of differential locus of control orientation, internal and external subjects are different on a number of behaviors. It is possible that they may also differ in how they respond to a test, internal subjects being more consistent or less aberrant and external subjects more aberrant or less consistent. Studies relating locus of control to risk taking behaviors have produced convincing evidence for the advancement of the proposition that locus of control and response pattern are associated. When playing a dice game, external subjects have the tendency of choosing risky alternatives, but without much confidence. The inconsistency demonstrated in their betting practice reveals that external subjects behaved in a confused manner. Furthermore, the above proposition is consolidated by the finding from studies concerning decision-making approaches of the internal and external subjects. Again as a result of disbelief in personal effort, external subjects treated all items the same way in terms of the time spent regardless of the wide range of difficulty levels. To make a difficult choice without extra effort and extra time suggests that external subjects yield to demanding situations easily and act in a random manner. This explanation is supported by the finding from another study that external subjects when faced with challenge are more intense with low tolerance to frustration. All this evidence convincingly lead to the expectation that locus of control and person reliability or response pattern are related at least to some extent.

Modification of Locus of Control

The discussion in the preceding section argues that internal-external control should be a useful construct for the understanding of person reliability and hypothesizes that locus of control and person reliability are related. If test takers with a low person reliability or aberrant response patterns do have an external control orientation and if, indeed, a relationship between person reliability and locus of control can be discovered, then it would appear logical to concentrate on the construct of locus of control of the test takers at this stage. That is, it would be important to provide evidence that perceived locus of control actually influences test taking behavior. Therefore, a starting point should be the manipulation of specific task expectancy so that a person's locus of control can be modified.

In 1957, Phares conducted his classic study to test the hypothesis that persons in a skill situation should use their past performance as a basis for generalizing about their future performance. To test this hypothesis, Phares had subjects perform the task under two conditions. Half the subjects received instructions that presented the tasks as skill types; the other half received instructions that encouraged a chance orientation. Reinforcement was controlled by the experimenter and was presented in a prearranged sequence. Before each trial, the subject had to bet whether he would perform the task correctly, thus providing a measure of expectancy. The results were in accord with the hypothesis. The skill instructions produce greater expectancy changes than the chance instructions. Subjects also shift their expectancies more often under the skill conditions.

The above research illustrates two points. First, the study shows that expectancies of persons in learning situations differ, depending upon whether the conditions are seen as involving skill or chance. Second, the orientation of the subjects was manipulated either through the use of instructions or the type of task presented. The possibility of modifying person's locus of control orientation allows researchers to examine the nature of the relationship between locus of control and person reliability.

The Present Study

The present study has two purposes. First, it intends to determine the relationship between test takers' person reliability or response pattern indices (the Modified Caution Index, MCI; the Person Average R, PAR; and the Maximum Log Likelihood Function, lo) and their locus of control orientation (LOC). Second, it attempts to examine the effect of the test instructions on test takers' locus of control orientation. It is expected that a shift in locus of control orientation will influence their person reliability and response pattern.

Major Hypothesis

1. H_1:

Subjects' locus of control (LOC) will correlate positively with their Modified Caution Index (MCI) and correlate negatively with their Person Average R (PAR) and their Maximum Log Likelihood Function (lo). That is subjects with an external locus of control orientation (high LOC scores) will have a more aberrant response pattern (high MCI, low PAR, and low lo scores).

2. H_1:

Subjects who receive external test instructions will have a more external locus of control orientation (high LOC scores) and have a more aberrant response pattern (high MCI, low PAR, and low lo scores) than those in the control group who receive neutral test instruction.

3. H_1:

Subjects who receive internal test instruction will have a more internal locus of control orientation (low LOC scores) and have a more consistent response pattern (low MCI, high PAR, and high lo scores) than those in the control group.

Minor Hypotheses

1. H_1:

Subjects' Modified Caution Index (MCI) will correlate negatively with their Person Average R (PAR) and their Maximum Log Likelihood Function (lo). That is subjects with a more aberrant response pattern will have high MCI (an indicator of more aberrancy), low PAR, and low lo (indicators of more consistency).

2. H_1:

Subjects' mathematics test scores (MT) will correlate negatively with their locus of control orientation (LOC) and Modified Caution Index (MCI), and correlate positively with their Person Average R (PAR) and their Maximum Log Likelihood Function (lo). That is, subjects with higher mathematics test scores (high MT scores) will have a more internal locus of control orientation (low LOC scores) and a more consistent response pattern (low MCI, high PAR, and high lo scores).

3. H_1:

Subjects' perception of a mathematics test's difficulty level will be dependent on the kind of test instructions they will receive. That is, in the external group there will be a larger proportion of subjects who believe that a mathematics test is difficult than those in the control group. In the internal group there will be a smaller proportion of subjects who believe that a mathematics test is difficult than in the control group.

4. H_1:

Subjects' expectancy of a mathematics test's outcome will be dependent on the kind of test instructions they will receive. That is in the external group there will be a larger proportion of subjects who believe that their chance will determine the outcome of a test than in the control group. In the internal group there will be a smaller proportion of subjects who believe that their chance will determine the outcome of a test than in the control group.

5. H_1:

Subjects' use of guessing test strategy will be dependent on the kind of test instruction they will receive. In the external group there will be a larger proportion of subjects who will use guessing test strategy than in the control group. In the internal group there will be a smaller proportion of subjects who will use guessing test strategy than in the control group.

6. H_1:

Subjects' attitude toward a mathematics test will be dependent on the kind of test instruction they will receive. That is in the internal group there will be a larger proportion of subjects who will take a mathematics test seriously than in the control group. In the external group there will be a smaller proportion of subjects who will take a mathematics test seriously than in the control group.

CHAPTER 2

METHOD

Subjects

The subjects of this study were 576 Chinese first-year senior middle school students of Changning District, which is one of the 12 districts under the municipality of Shanghai, a seaport on the east coast of China. First-year students of senior middle school in China are equivalent to tenth-grade students in the United States, both in terms of their age (around 16 years old) and schooling (10 years). Schools in the 12 Shanghai districts adopted a uniform curriculum and the same textbooks compiled and recommended by the Municipal Bureau of Education. These 576 students came from 12 classes, 48 from each class. Of the 576 students, 192 were from four classes, the entire first-year senior middle school student body at Yanan middle School. The rest of the 384 students were from eight classes of eight middle schools in the same district. The names of these eight middle schools were Xingwu Middle School, Tianshan No.2 Middle School, Fanyu Middle School, Shanghai No.3 Girls Middle School, Yuping No.1 Middle School, Yuping

No.2 Middle School, Fudan Middle School, and Jianqing Middle School.

The sex ratio of this subject pool approximated one to one with 283 students being male and 293 female. In 11 classes, however, there were more male students than female students, as is the case in most of the classes at this level in China. This sex ratio changed because of the inclusion of a class of female students from a girls middle school which was not common in China even in the cities. The average class size in China is around 50 students in most of the middle schools in Chinese cities at that time. In the present study 48 students were sampled from each class by including only the first 48 numbered test instruments from the total number of test instruments distributed.

Materials

The materials employed in this study included four sections. They were the Test Instruction, Mathematics Test, Rotter's Internal-External Control Scale, and Survey Questionnaire.

<u>1. Test Instruction</u>

Test instruction is the key manipulated variable to this study in its effort to show the possible effect of locus of control on response pattern and person reliability. This part included three different instructions: internal, external, and control test instructions. The three instructions were designed to have positive, negative, and neutral effect on subjects' perception of control and were further expected that a shift in LOC would influence their test-taking behavior which would be reflected in their person reliability indices.

<u>a. Internal Instruction</u>.

The internal instruction was written with the purpose of enhancing the students' expectancy of positive test results, changing their locus of control orientation in the direction of internality, and producing a more consistent response pattern. It reiterated that past experience (the test results of students who had taken this test previously) had proved that the test was within the range of his/her academic ability, and emphasized that if he/she mobilized his/her knowledge and skill and tried hard, he/she could achieve satisfactory results. The essential message which this instruction attempted to convey was positive, supportive and encouraging. The literal English translation from the Chinese version for internal instruction is presented below:

> "We have given this test to a large number of students just like you. We know that this is a difficult test but within tenth grade capabilities. You might find many of the questions are answerable. This means, if you think things through and try very hard, you will be able to answer many of the questions correctly. Although we are quite sure that people are able to answer many of the questions correctly, we need a few more people to prove it statistically.
>
> Important! Answer all questions."

b. External Instruction.

Opposite to the internal instruction, the external instruction was intended to sway the subjects' minds to the external side of the locus of control continuum, result in a less consistent response pattern. The message sent out from this instruction warned the subjects that no matter how hard one would try, the outcome would always be the same, i. e., disappointing because of the difficulty level of the test. It implied that the subjects would think that whether one answered the items right or wrong largely depended on luck, and that the degree of one's effort made little difference. The literal English translation of the Chinese version of the external instruction is presented below:

> "We have given this test to a large number of students just like you. We know that this is a difficult test beyond the tenth grade capacities. You might find many of the questions are too difficult. This means, if you think things through and try very hard, you will still be unable to answer many of the questions correctly. Although we are quite sure that people are unable to answer many of the questions correctly, we need a few more people to prove it statistically.
>
> Important! Answer all questions."

c. Control Instruction.

The third instruction for the control group made no attempt to produce any particular change in the subjects' expectancy of the test outcome, locus of control orientation, and their response pattern. This instruction was parallel to the other two instructions in format, but with an intended neutral message, neither encouraging nor depressing. The literal English translation of the Chinese version for the control instructional treatment is presented below:

> "We have given this test to a large number of students just like you. We know that this is a test of average difficulty level. However, some students have done very well, answering many of the questions correctly, while some students have done rather poorly, missing many of the questions. Since we want to make sure how difficult this test is, we need a few more students to prove it statistically.
>
> Important! Answer all questions."

2. Mathematics Test

The subjects' responses to the mathematics test items served as the basis for calculating the three person reliability or response pattern indices, MCI, PAR, lo, and MT. In order to detect the change of students' person reliability effectively, an ideal test instrument as suggested by Rotter and other researchers should meet an important requirement, "appropriate" difficulty level. This study entailed the use of a test relatively difficult for the target subjects. If a test were too easy, then the majority of the students would get most of the items correct. On the other hand, if it were too difficult, the majority of the students could fail most of the items. Either way would limit the number of effective items. A test, too easy or too difficult, would restrict the range of subjects' person reliability or response pattern indices. As a result, it would be more difficult to identify the relationship between person reliability and locus of control, and to detect the effect of the test instructions on the three variables, response pattern indices (MCI, PAR, and lo), and locus of control (LOC), and a mathematics test (MT).

The mathematics test of this study consisted of two parts: the quantitative part of the Graduate Record

Examination (GRE)-like test and a section selected from a MENSA-like sample test. MENSA is the international organization whose only requirement for membership is an I.Q. in the "genius" range. The GRE test is a standardized test designed to examine the aptitude of the applicants for graduate programs in the United States. The majority of the Chinese graduate students who had taken the GRE over the past few years agreed that the quantitative part of the GRE test was relatively easy for them and that an average Chinese middle school graduate should be able to take that part of the GRE satisfactorily. Also, the quantitative part of the GRE test, compared with verbal and reasoning part, is less culturally loaded. The instructions are short and simple and less difficult to translate from English into Chinese. The most important reason for selecting the GRE-like test was that, to my knowledge at that point, there was no standardized test similar to the GRE-like test available in China. The GRE-like part of the mathematics test contains 30 items (the first 30 items of the mathematics test). (See appendix B)

For the same reason presented above, the second part of the mathematics test was selected from a MENSA-like test, a test with a strong emphasis on perception of figures. The differences between the GRE-like and the MENSA-like parts of the mathematics test was in the design of the items and the skills required to solve the problems. The MENSA-like part of the mathematics test consisted of 10 items (the last 10 items of the mathematics test). (See Appendix B)

3. Rotter's Internal-External Control Scale

The Rotter's (1966) Internal-External Control Scale (See Appendix C) was chosen as the measure of subjects' locus of control orientation from several paper and pencil measures of generalized locus of control. The basis for Rotter's internal and external locus of control measure was first developed by Phares (1957) and James (1957) who found that patterns of acquisition and extinction of reinforcement under skill and chance conditions differed from "internal" and "external" subjects. Their measure then was broadened with additional items and analyzed across a large number of subjects by Rotter (1966), resulting in an IE scale which is the choice for this study.

Measurement of how people perceive their ability to control events or outcomes in their environment is determined by the total score. The instrument, forced-choiced in design, contained 29 items, with six fillers. Scores can range from 0 to 23, with each response counting one point. Persons of the sample are identified as either having an more internal or external locus of control orientation with the mean of the sample group as the cut-off point. According to Rotter, internal consistency estimates are between .65 and .79, and test-retest reliability is reported to vary from .49 to .83 depending on the type of sample and the time interval involved. Some others (Hersch & Scheible, 1967; Harrow & Ferrante, 1969) found reliabilities ranging from .48 to .84.

4. Survey Questionnaire

In order to study the effectiveness of the three instructional treatments internal/skill, external/chance, and neutral/control on the subjects' locus of control orientation and test-taking behavior represented by person reliability indices (MCI, PAR, and lo), eight follow-up questions were created. They were designed to study the subjects' perception of the test difficulty level, test result expectancy, test strategy, and others. Subjects' answers to these questions would provide additional information for the interpretation and discussion of the findings. The literal English translation of the Chinese version of the Survey Questionnaire is presented below:

Survey Questions

Please read the following questions carefully, then circle the responses or fill in the numbers you feel most appropriate.

1. Do you think the test was difficult?	Yes	No
2. Do you think the outcome of this test was dependent on your...	Ability	Luck
3. Did you guess in taking today's test?	Yes	No
4. Did you try your best in today's test?	Yes	No
5. Your sex is	M	F

All the materials were translated into Chinese by way of back-translation (Brislin, 1980). Three bilingual persons fluent in both Chinese and English were invited as the consultants of this study. First, one of them translated the English version of the materials into Chinese, then another person translated the Chinese version back into English. The third person compared the new English version with the original English version to find any serious discrepancies between the two. Finally, three of them worked together to decide on the best possible Chinese version of the whole package. The purpose of this time-consuming practice was to ascertain that the two versions were as identical as possible, in order to reduce the discrepancy due to translation from English into Chinese to a minimum. They also made some modifications in the choice of the words so that Chinese students could understand the statements.

The above four parts of the materials included in the package followed this sequence: Test Instruction, Mathematics Test, Rotter's Internal-External Locus of Control Scale, and Survey Questionnaire. Efforts were made to guarantee that each group had an equal number of subjects. The three-way factorial design (treatment by sex by class) of this study is presented in Table 5.

Procedures

The experimental session, which was conducted in the classroom setting, lasted approximately one and half hours. The whole session could be divided into three stages. First, subjects were randomly assigned to an instructional group by receiving one of the three packages. The students were asked to read the instruction carefully. Then, the subjects were allowed to proceed to take the mathematics test. After the subjects had finished the mathematics test, the subjects completed work on Rotter's Internal-External Control Scale. In the final stage, the subjects answered the survey questions. The subjects were then debriefed, thanked and dismissed. After the classroom session a small number of

Table 2

The Three Way Design (Class by Sex by Treatment) of the Present Study

Class	Sex	Treatment External	 Internal	 Control
1	M F	16	16	16
2	M F	16	16	16
3	M F	16	16	16
4	M F	16	16	16
5	M F	16	16	16
6	M F	16	16	16
7	M F	16	16	16
8	M F	16	16	16
9	M F	16	16	16
10	M F	16	16	16
11	M F	16	16	16
12	M F	16	16	16
		192	192	192

the subjects were interviewed individually for more direct and detailed feedback about the whole experiment.

Since the average middle school class size in that school district at that time was around 50 students, extra copies of the packages were prepared to ensure every student in the class had a copy to work on, but only the first 48 numbered copies were included in the experiment.

Data Analysis

The overall data set included three subsets, the first one, a person item response matrix from the mathematics test; the second, a person item response matrix from Rotter's Internal-External Control Scale; and the third, a person item response matrix from the survey questionnaire items.

The person-item response matrix from the mathematics test was used to compute the three person reliability indices, PAR, MCI and lo. MCI was computed on a microcomputer using spread sheets prepared for the MCI formula. Both mainframe and PC, were used to compute the Maximum Log Likelihood Function (lo). The three parameters of the mathematics test based on the IRT model, item discrimination (a), item difficulty (b), pseudo-chance level (c), and fixed ability theta (θ) were first computed on the mainframe computer by using LOGIST. These four numbers and the person item response matrix on the mathematics test were then applied to the lo formula prepared on PC spread sheets.

The three obtained person reliability indices, MCI, PAR and lo, together with the measure of locus of control, LOC, the scores on a Math Test, MT, and the data from the survey questionnaire, formed a combined dataset as the input for hypothesis testing. The dataset was first used to compute the correlation matrix of the five variables, MCI, PAR, lo, LOC, and MT. The combined data set was then utilized to determine whether the independent variables, treatment (test instructions) and sex, had any effect on the five dependent variables, first, LOC, then, MCI, PAR, lo, and MT by using analysis of variance (ANOVA) procedure, with means and Scheffé test as options for the purpose of identifying the pairs with a difference statistically significant at the .05 level. Lastly, the Chi-square test was performed on the combined dataset to examine the effect of the treatment on the subjects' response to the survey questionnaire items, their perception of test difficulty level, test outcome expectancy, the use of guessing test-taking strategy, and their attitude toward the test.

CHAPTER 3

RESULTS

This chapter includes four sections: first, descriptive statistics on the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), locus of control (LOC), and a mathematics test (MT); second, the relationship between MCI, PAR, lo, LOC, and MT; third, the effect of the treatment on LOC, MCI, PAR, lo, and MT; and fourth, the results of the Chi-square tests on the data from the questionnaire items.

Descriptive Statistics on MCI, PAR, lo, LOC, and MT

The mean, maximum score, minimum score, and standard deviation of the five variables, MCI, PAR, lo, LOC, and MT are presented in Table 3. The mean of MCI was 0.20 with a standard deviation of .09 and a range from .47 to .00. The obtained statistics were expected. The theoretical upper and lower limit of MCI is one and zero respectively. A larger MCI means a more errorful response pattern. The mean of PAR was .24 with a range from .39 to .04, and a standard deviation of .06. These statistics were within the expected boundary. The expected range of PAR extends from one to zero. In contrast to MCI, a smaller PAR stands for a more aberrant response pattern. In other words, a larger PAR represents a more consistent response pattern. Always a negative number, lo had a mean of -14.46, a range from -20.79 to -4.96, and a standard deviation of 2.52. The lo values reported in different studies varied, sometimes substantially. Compared with other studies, the mean and the range of lo in the present study were slightly larger than the average of the lo's reported so far. Like PAR, a small lo stands for a more aberrant response pattern or low person reliability. LOC representing a person's internal-external locus of control orientation, ranged from 2 to 20 with a mean of 11.54 and a standard deviation of 3.11. The mean is slightly larger than the means computed based on the data in most of the studies. A LOC score can be as high as 23 and as low as zero. A high LOC indicates that this person tends to believe more strongly that luck and chance are in control of the outcome of an event. MT (score on the mathematics test) ranges from 10 to 37 with a mean of 25.93 and a standard deviation of 4.52. The mathematics test contained 40 items which means the highest possible score can be 40. No one in the study achieved a perfect score of 40. The highest score was 37. The outcomes were expected since it was intended to be a fairly difficult test for the target subjects.

The frequency distributions of MCI, PAR, lo, LOC, and MT are graphically presented in Figures 3 to 7.

Table 3

Summary Statistics of the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), Locus of Control (LOC), and a Mathematics Test (MT)

	N	Mean	Std Dev	Minimum	Maximum
MCI	576	.20	.09	.00	.47
PAR	576	.24	.06	.04	.39
lo	576	-14.46	2.52	-4.96	-20.79
LOC	576	11.54	3.11	2	20
MT	576	25.93	4.52	10	37

```
MIDPOINT
  MCI                                FREQ  CUM.     PERCENT    CUM.
                                           FREQ                PERCENT
          |
0.000     |*                            3     3        0.52       0.52
0.025     |*                            1     4        0.17       0.69
0.050     |***                         13    17        2.26       2.95
0.075     |******                      30    47        5.21       8.16
0.100     |*******                     36    83        6.25      14.41
0.125     |*********                   46   129        7.99      22.40
0.150     |****************            78   207       13.54      35.94
0.175     |**************              69   276       11.98      47.92
0.200     |**********                  51   327        8.85      56.77
0.225     |**********                  52   379        9.03      65.80
0.250     |**********                  48   427        8.33      74.13
0.275     |********                    42   469        7.29      81.42
0.300     |********                    41   510        7.12      88.54
0.325     |*****                       27   537        4.69      93.23
0.350     |***                         16   553        2.78      96.01
0.375     |**                          12   565        2.08      98.09
0.400     |*                            4   569        0.69      98.78
0.425     |*                            6   575        1.04      99.83
0.450     |                             0   575        0.00      99.83
0.475     |*                            1   576        0.17     100.00
          ----+---+---+---+
             20  40  60  80

             FREQUENCY
```

Figure 3

Frequency Distribution of the Modified Caution Index (MCI)

```
MIDPOINT
  PAR                                       FREQ  CUM.   PERCENT    CUM.
                                                  FREQ            PERCENT
          |
0.025     |                                   0      0     0.00       0.00
0.050     |*                                  3      3     0.52       0.52
0.075     |*                                  7     10     1.22       1.74
0.100     |**                                10     20     1.74       3.47
0.125     |*****                             23     43     3.99       7.47
0.150     |********                          41     84     7.12      14.58
0.175     |***********                       53    137     9.20      23.78
0.200     |*************                     64    201    11.11      34.90
0.225     |*************                     65    266    11.28      46.18
0.250     |****************                  82    348    14.24      60.42
0.275     |******************** 102                450    17.71      78.13
0.300     |*************                     65    515    11.28      89.41
0.325     |********                          42    557     7.29      96.70
0.350     |***                               16    573     2.78      99.48
0.375     |*                                  3    576     0.52     100.00
0.400     |                                   0    576     0.00     100.00
          ----+---+---+---+---+
             20  40  60  80  100

                FREQUENCY
```

Figure 4

Frequency Distribution of the Person Average R (PAR)

MIDPOINT lo		FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
-20	*	5	5	0.87	0.87
-19	****	21	26	3.65	4.51
-18	*********	46	72	7.99	12.50
-17	**********	52	124	9.03	21.53
-16	*******************	96	220	16.67	38.19
-15	**************	70	290	12.15	50.35
-14	*****************	85	375	14.76	65.10
-13	*************	67	442	11.63	76.74
-12	*************	66	508	11.46	88.19
-11	*******	33	541	5.73	93.92
-10	****	19	560	3.30	97.22
-9	**	9	569	1.56	98.78
-8	*	3	572	0.52	99.31
-7	*	1	573	0.17	99.48
-6	*	2	575	0.35	99.83
-5	*	1	576	0.17	100.00

```
----+---+---+---+---
   20  40  60  80

      FREQUENCY
```

Figure 5

Frequency Distribution
of the Maximum Log Likelihood Function (lo)

MIDPOINT LOC		FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
2	*	1	1	0.17	0.17
3		0	1	0.00	0.17
4	***	7	8	1.22	1.39
5	****	11	19	1.91	3.30
6	****	9	28	1.56	4.86
7	********	20	48	3.47	8.33
8	******************	45	93	7.81	16.15
9	************************	59	152	10.24	26.39
10	**************************	66	218	11.46	37.85
11	*****************************	72	290	12.50	50.35
12	***************************	68	358	11.81	62.15
13	**************************	64	422	11.11	73.26
14	********************	51	473	8.85	82.12
15	******************	44	517	7.64	89.76
16	**********	24	541	4.17	93.92
17	********	21	562	3.65	97.57
18	***	7	569	1.22	98.78
19	**	6	575	1.04	99.83
20	*	1	576	0.17	100.00

```
----+----+----+----+----+----+----+-
   10   20   30   40   50   60   70
```

FREQUENCY

Figure 6

Frequency Distribution of Locus of Control (LOC)

MIDPOINT MT		FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
10	*	1	1	0.17	0.17
11		0	1	0.00	0.17
12		0	1	0.00	0.17
13	**	3	4	0.52	0.69
14	**	4	8	0.69	1.39
15	*	2	10	0.35	1.74
16	**	3	13	0.52	2.26
17	*****	9	22	1.56	3.82
18	*******	13	35	2.26	6.08
19	********	15	50	2.60	8.68
20	*************	26	76	4.51	13.19
21	*************	26	102	4.51	17.71
22	**************	27	129	4.69	22.40
23	*****************	34	163	5.90	28.30
24	*********************	41	204	7.12	35.42
25	*********************	41	245	7.12	42.53
26	*************************	49	294	8.51	51.04
27	*****************************	57	351	9.90	60.94
28	************************	48	399	8.33	69.27
29	************************	47	446	8.16	77.43
30	**********************	44	490	7.64	85.07
31	*************	26	516	4.51	89.58
32	****************	31	547	5.38	94.97
33	*****	10	557	1.74	96.70
34	*******	13	570	2.26	98.96
35	**	3	573	0.52	99.48
36	*	2	575	0.35	99.83
37	*	1	576	0.17	100.00

-----+----+----+----+----+----

10 20 30 40 50

FREQUENCY

Figure 7

Frequency Distribution of a Mathematics Test (MT)

The Relationships Between MCI, PAR, lo, LOC, and MT

The Relationships Between MCI, PAR, and lo

The matrix of correlations between MCI, PAR, lo, LOC and MT is reported in Table 4.1. All three correlations between the three response pattern indices (MCI, PAR, and lo) were significant at the 0.005 level or better. The correlation between MCI and PAR ($r=-.95$, $df=574$, $p<.0001$) was very strong. Less strong are the correlation between MCI and lo ($r=-.42$, $df=574$, $p<.0001$) and between PAR and lo ($r=.46$, $df=574$, $p<.0001$). The results indicate that a high MCI goes with a low PAR and low lo, or the higher one's PAR and lo, the lower one's MCI will be. To put it differently, a high person reliability, or a response pattern with an indicator of high consistency will naturally yield a small indicator of aberrance.

The Relationships Between MCI, PAR, lo, and LOC

All three response pattern indices, MCI, PAR and lo, had a statistically significant relationship with locus of control, LOC at the 0.005 level or better, though different in degree and opposite in directions (the correlation between MCI and LOC: $r=.61$, $df=574$, $P<.0001$; the correlation between PAR and LOC: r=-.60, df=574, P<.0001; the correlation between lo and LOC: r=-.12, df=574, p<.005). It means that a person with a high locus of control score (more external orientation) tended to have a high MCI (more aberrant), and a low PAR and lo (less reliable), thus a person with a more external locus of control orientation control orientation tends to have a low person reliability or a more aberrant or less consistent response pattern. The relationship between lo and LOC was not as strong as the one between MCI and LOC and between PAR and LOC.

Table 4.1

Pearson Product-Moment Correlation Coefficients Between Five Variables: the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), Locus of Control (LOC), and a Mathematics Test (MT)

	MCI	PAR	lo	LOC	MT
MCI	1.00	-.95**	-.42**	.61**	-.26**
PAR		1.00	.46**	-.60**	.22**
lo			1.00	-.12*	.19**
LOC				1.00	-.52**
MT					1.00

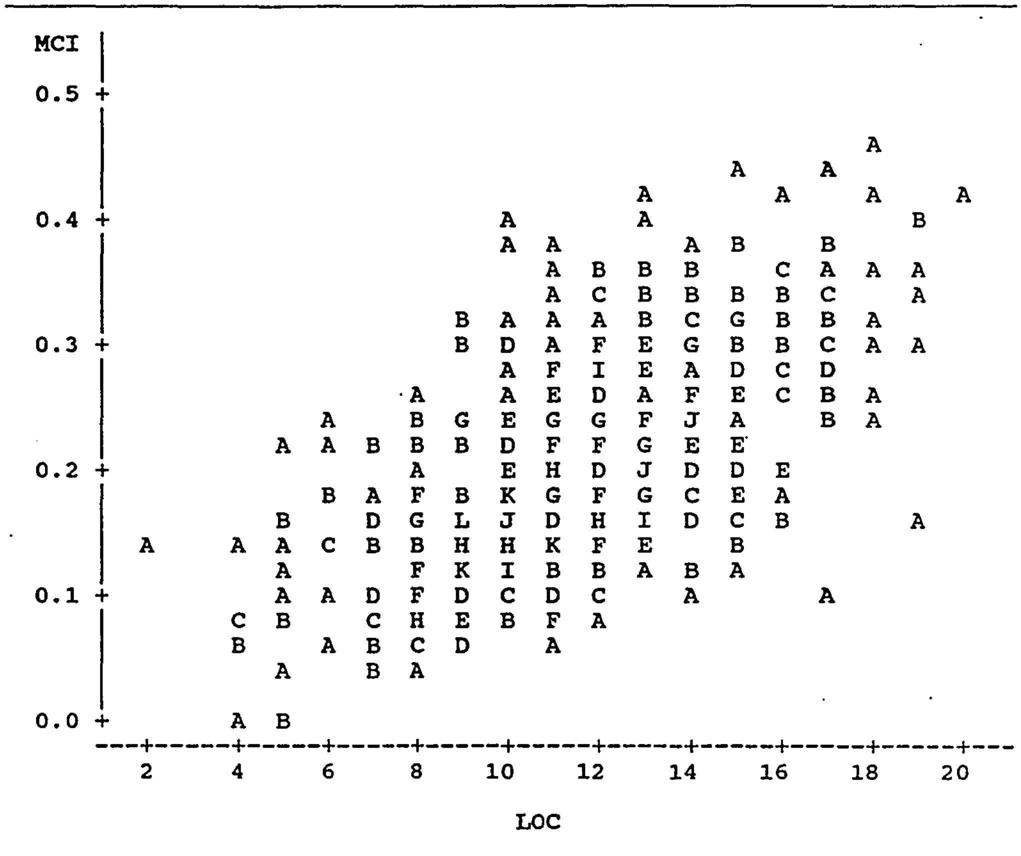
* p<.005
** p<.0001
df=574

The Relationships Between MCI, PAR, lo, LOC, and MT

The relationships between the three response pattern indices (MCI, PAR, and lo) and the mathematics test score (MT) were low but statistically significant at the 0.0001 level (the correlation between MCI and MT: r=-.26, df=574, p<.0001; the correlation between PAR and MT: r=.22, df=574, p<.0001; the correlation between lo and MT: r=.19, df=574, p<.0001)). A person with a high mathematics test score tended to have a high PAR and lo, and a low MCI, indicating a high person reliability or a normal response pattern.

LOC and MT were negatively correlated (r=-.52, df=574, p<.0001), which suggests that a person with a more external locus of control orientation, a high LOC, seemed to score low on the mathematics test. All the correlations in the presented matrix, though with a sizable range from .95 to 0.12 in absolute value, were significant at the .005 level or better with a degree of freedom of 574.

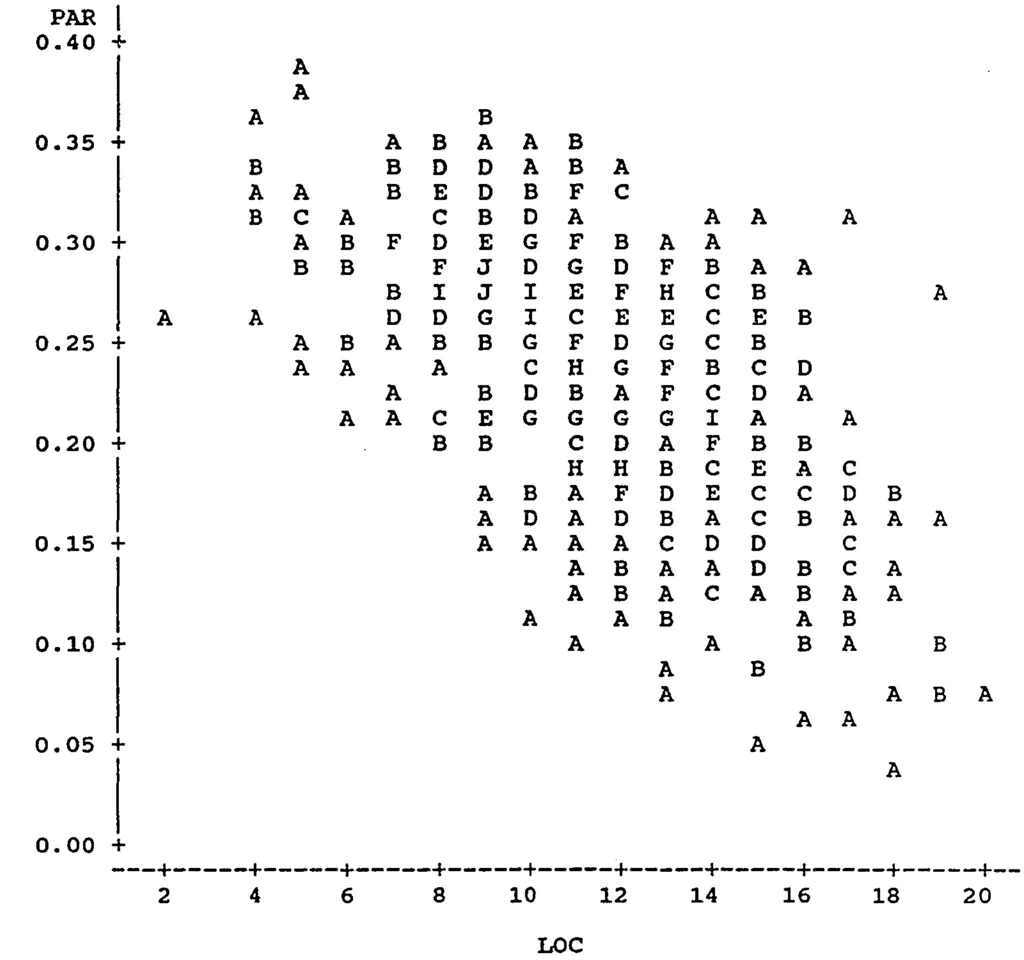
As can be seen in Figure 8 the relationship between MCI and LOC is most likely a linear one, as is the relationship between PAR and LOC (see Figure 9). The type of relationship between lo and LOC was not as clearly discernible as demonstrated in Figure 10.



LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Figure 8

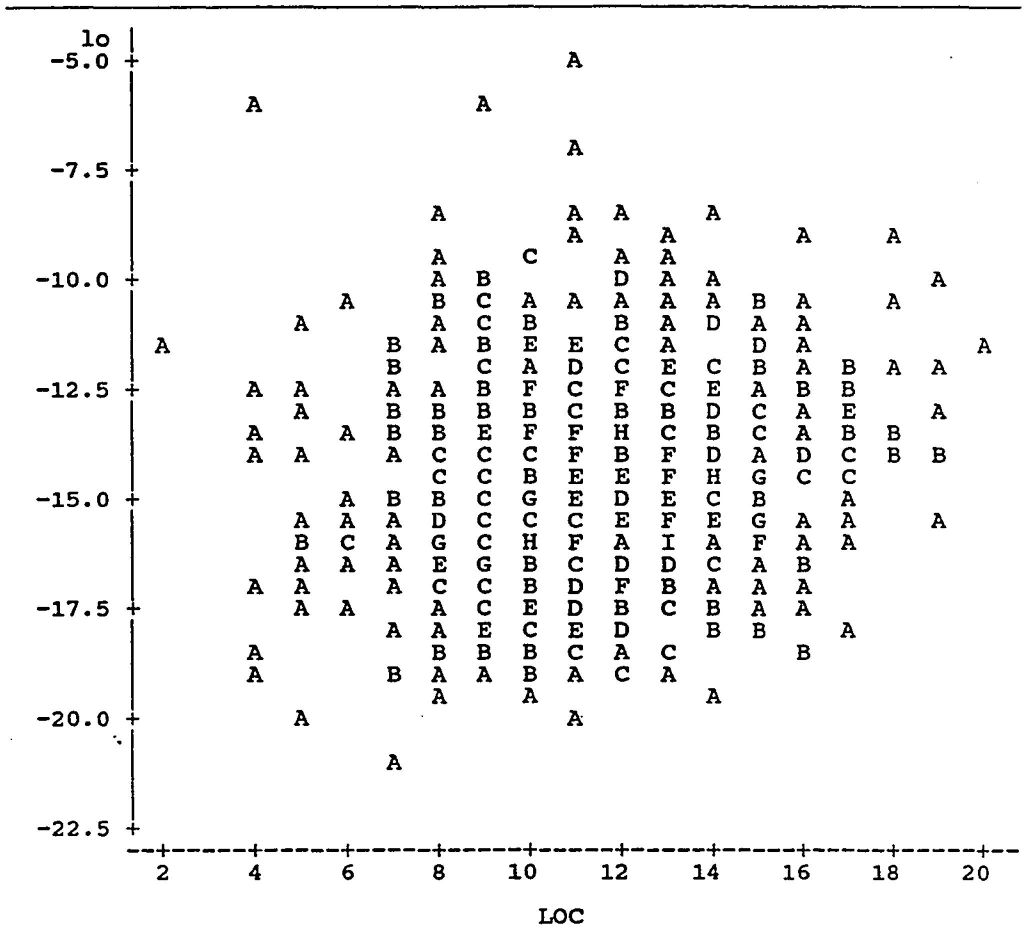
Bivariate Plot of the Modified Caution Index (MCI)
with Locus of Control (LOC)



LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Figure 9

Bivariate Plot of the Person Average R (PAR) with the Locus of Control (LOC)



LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Figure 10

Bivariate Plot of the Maximum Log Likelihood Function (lo) with Locus of Control (LOC)

The Partial and Semi-Partial Correlations Between MCI, PAR, lo, and LOC, Controlling for MT

Since the three person reliability or response pattern indices (MCI, PAR, and lo) and LOC were all correlated with MT, this might, to some extent, contribute to the obtained strong relationship between MCI and LOC, and between PAR and LOC, and the much weaker but significant relationship between lo and LOC. To ascertain the strength of the relationships between the three response pattern indices (MCI, PAR, and lo) and LOC, which were not due to the contribution of MT, the partial correlation procedure was performed, controlling for MT from both the first variable, response pattern indices (PAR, MCI, and lo) and the second variable, LOC, respectively. Also computed was the semi-partial correlation controlling for MT only from the second variable, LOC. A comparison of the three correlations is shown in Table 4.2.

As expected, the results indicate that the partial and semi-partial correlations between two response pattern indices (MCI and PAR) and locus of control (LOC) were smaller than the original simple correlations; but the differences were almost minimal. This means that the relationship between response pattern indices (MCI and PAR) and LOC remains strong when partialing out the effect of MT

Table 4.2

Comparison of Three Correlation Coefficients: Pearson Product-Moment Correlation, Partial Correlation (Controlling for a Mathematics Test, MT, from both Variables), and Semi-Partial Correlations (Controlling a Mathematics Test, MT, from LOC) Between the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), and Locus of Control (LOC)

	Pearson Correlation	Partial Correlation (MT Partialed out from Both Variables)	Semi-partial Correlation (MT partialed out from LOC)
MCI/LOC	.61***	.58***	.56***
PAR/LOC	-.60***	-.58***	-.56***
lo/LOC	-.12**	-.04	-.03

** p<.005
*** p<.0001
df=576

from both variables of the relationship (person reliability indices: MCI and PAR; and LOC) or just one variable LOC. A person's score on a mathematical test imposes little effect on the relationship between person reliability (MCI and PAR) and locus of control orientation, LOC. With regard to the correlation between lo and LOC, since the original correlation was already quite small though still significant, when the effect of MT was partialed and semipartialed out, the relationship between lo and LOC was not significant at the 0.05 level (partial correlation between lo and LOC controlling for MT from both variables: r=.04, df=574, p>.05; the semi-partial correlation between lo and LOC controlling for MT from LOC: r=.03, df=574, p>.05).

Because the correlation based on the data from the total group might be affected by the treatment, the correlation analysis was further carried on with the data solely from the control group which was not affected by the treatment, i.e., either internal or external test instructions. As shown in Tables 5.1-5.2, the simple

Table 5.1

Pearson Product-Moment Correlation Coefficients
Between Five Variables:
the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), Locus of Control (LOC), and a Mathematics Test (MT) Based on the Control Group Data

	MCI	PAR	lo	LOC	MT
MCI	1.00	-.96**	-.40**	.66**	-.23**
PAR		1.00	.43**	-.66**	.19**
lo			1.00	-.11*	.29**
LOC				1.00	-.55**
MT					1.00

* p<.005
** p<.0001
df=574

partial, and semi-partial correlations between different pairs of variables are almost the same as in the total group with minimal change.

Table 5.2

Comparisons of Three Correlation Coefficients: Pearson Product-Moment Correlation, Partial Correlation (Controlling for a Mathematics Test, MT, from both Variables), and Semi-Partial Correlations (Controlling a Mathematics Test, MT, From LOC) Between the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), and Locus of Control (LOC) Based on the Control Group Data

	Pearson Correlation	Partial Correlation (MT Partialed out from Both Variables)	Semi-partial Correlation (MT partialed out from LOC)
MCI/LOC	.66***	.65***	.55***
PAR/LOC	-.68***	-.67***	-.56***
lo/LOC	-.11**	.06	.05

** p<.005
*** p<.0001
df=576

The Effect of the Treatment on LOC, MCI, PAR, Lo, and MT

Results of Analysis of Variance

Analysis of variance (ANOVA) was used to test the hypotheses associated with the proposed causal relationship between locus of control and person reliability or response pattern indices, MCI, PAR, and lo, through the manipulation of the treatment which consisted of three different test instructions, internal, external, and control. First, the effect of the treatment was tested together with two organismic variables contained in the test design sex and class to ascertain their major effects or joint effects upon the four dependent variables, the Modified Caution Index (MCI), the Person Average Reliability (PAR), the Maximum Log Likelihood Function (lo), and locus of control (LOC). Class was the smallest unit of analysis of the study. A total of 576 subjects consisted of 12 classes from eight schools, 48 from each class. The results of ANOVA indicate that independent variables treatment and class had statistically significant effect on the dependent variables, each of which accounted for a statistically significant amount of variance of all four independent variables, LOC, MCI, PAR, and lo. However, according to the Scheffé test, the means of LOC, MCI, PAR, and lo of the 12 classes were not significantly different from each other. Sex was not a significant factor in determining the value of any of the four dependent variables. There were not any significant interactions between treatment, class, and sex. Therefore, the variables sex and class were excluded from the following ANOVA, and only the results of a one-way ANOVA with the treatment as the sole independent variable are reported here. All four ANOVA achieved statistically significant test results indicating that the treatment or test instruction was an effective factor in determining the value of the dependent variables of LOC, MCI, PAR, and lo. The Scheffe test following each of the ANOVA produced largely consistent results only with some slight differences. The detailed description of the ANOVA and Scheffe test results are now presented below.

The ANOVA on the LOC data resulted in a significant main effect for treatment ($F=13.20$, $df=2; 573$, $p<.001$). The results of this analysis are reported in Table 6.1. The results of the Scheffe test (Table 6.2) indicate that the mean LOC for the three treatment groups were divided into two blocks with mean LOC's significantly different at the 0.05 level. The first block included one group, the external group (mean=12.43). The second block included two groups, the control group (mean=11.30) and the internal group (mean=10.89) which were not significantly different. This means that subjects who received external test instructions

Table 6.1

The Analysis of Variance for the Effect of the Treatment on Locus of Control (LOC)

Source	SS	DF	MS	F	P
Treatment	244.6979	2	122.3489	13.20	249E-8
Error	5312.5364	573	9.2714		
Total	5557.2343	575			

Table 6.2

The Scheffé Test for the Effect of the Treatment on Locus of Control (LOC)

Scheffé Grouping	Mean	N	Treatment
A	12.43	192	External
B	11.30	192	Control
B	10.89	192	Internal

Means with the same letter are not statistically significantly different ($p<.05$).

tended to have a more external locus of control orientation than those in the control group; on the other hand, those who received internal test instruction appeared to have a more internal locus of control oriented than those in the control group, however the difference was not large enough to warrant the claim of achieving statistical significance at the .05 level.

The results of the one-way ANOVA on MCI data are reported in Table 7.1. The analysis yielded significant MCI difference due to the treatment ($F=18.64$, $df=2; 573$, $p<.001$). The results of the Scheffé test (Table 7.2) show that the mean MCI of the internal group which received internal test instruction (mean=.18) was the smallest, significantly different from the mean MCI of the control group (mean=.21). The mean MCI of the external group which received external test instruction (mean=.23) was not significantly different from, or larger than the mean MCI of the control group (mean=.21). In other words, the response patterns of subjects who received an encouraging test instruction were less errorful than those of subjects from the control group; while the response patterns of those with negative test instruction were not necessarily more errorful, if measured by MCI.

Table 7.1

The Analysis of Variance for the Effect of the Treatment on the Modified Caution Index (MCI)

Source	SS	DF	MS	F	P
Treatment	.2556	2	.1278	18.64	14E-9
Error	3.9295	573	.0068		
Total	4.1851	575			

Table 7.2

The Scheffé Test for the Effect of the Treatment on the Modified Caution Index (MCI)

Scheffé Grouping	Mean	N	Treatment
A	.23	192	External
A	.21	192	Control
B	.18	192	Internal

Means with the same letter are not statistically significantly different ($p<.05$).

The results of a one-way ANOVA on PAR data are shown in Table 8.1. The independent variable, treatment, yielded a significant F value (F=20.19, df=2; 573, p<.001). The kind of treatment received determined the magnitude of PAR, how similar a person's response pattern was to that of his/her peers. The results of the Scheffé test presented in Table 8.2 show that the means of the three treatment groups were all statistically different from each other at the 0.05 level. The mean PAR of the external group which received the external test instruction (mean=.22) was the smallest; the mean PAR of the internal group which received the internal test instruction (mean=.26) was the largest. The mean PAR of the control group (mean=.23) fell somewhere in the middle between the mean PAR of the external group and the mean PAR of the internal group and with significant differences from both of them. Subjects who received internal test instructions responded to the test items more similar to that of their peers than those receiving external test instructions.

As in the cases of MCI and PAR, the ANOVA on the Maximum Log Likelihood Function, lo data obtained a statistically significant main effect for treatment (F=11.61, df=2; 573, p<.001). The results of this analysis are presented in Table 9.1. Further analysis by the Scheffé

Table 8.1

The Analysis of Variance for the Effect of the Treatment on the Person Average R (PAR)

Source	SS	DF	MS	F	P
Treatment	.1559	2	.0779	20.19	34E-10
Error	2.2131	573	.0038		
Total	2.3690				

Table 8.2

The Scheffé Test for the Effect of the Treatment on the Person Average R (PAR)

Scheffé Grouping	Mean	N	Treatment
A	.22	192	External
B	.23	192	Control
C	.26	192	Internal

Means with the same letter are not statistically significantly different ($p<.05$).

test (Table 9.2) indicates that the lo mean of the treatment group with external test instruction (mean=-15.13) was significantly different from, or smaller than, the lo mean of the control group (mean=-14.31) and the lo mean of the internal group (mean=-13.95). The lo mean of the latter two groups, the control and internal groups, were not significantly different at the 0.05 level. The results from this analysis mean that the negative test instruction produced response patterns contributing little to the maximum likelihood function while the positive test instruction failed to show any effect in the subject's response pattern, when measured by lo.

Contrary to the cases of LOC, MCI, PAR, and lo, the ANOVA on a mathematics test (MT) reports no statistically significant main effect for treatment (F=1.31, df=2; 573, p<.2700). The results of this analysis is presented in Table 10.1. Further analysis by the Scheffe test (Table 10.2) indicates that the differences between the MT mean of the external group (mean=25.49), the MT mean of the control group (mean=26.07), and the MT mean of the internal group (mean=26.19) were too small to claim statistical significance. The results indicate that the test instructions, internal or external, had no statistically significant effect on the subjects' mathematics test scores.

Table 9.1

The Analysis of Variance for the Effect of the Treatment on the Maximum Log Likelihood Function (lo)

Source	SS	DF	MS	F	P
Treatment	142.3341	2	71.1670	11.61	114E-7
Error	3511.1425	573	6.1276		
Total	3653.4767				

Table 9.2

The Scheffé Test for the Effect of the Treatment on the Maximum Log Likelihood Function (lo)

Scheffé Grouping	Mean	N	Treatment
B	-13.95	192	Internal
B	-14.31	192	Control
A	-15.13	192	External

Means with the same letter are not statistically significantly different ($p<.05$).

Table 10.1

The Analysis of Variance for the Effect of the Treatment on a Mathematics Test (MT)

Source	SS	DF	MS	F	P
Treatment	53.4826	2	26.7413	1.31	0.2700
Error	11676.8437	573	20.3784		
Total	11730.3263				

Table 10.2

The Scheffé Test for the Effect of the Treatment on a Mathematics Test (MT)

Scheffé Grouping	Mean	N	Treatment
A	26.19	192	Internal
A	26.07	192	Control
A	25.49	192	External

Means with the same letter are not statistically significantly different ($p<.05$).

The results of testing for the effect of the independent variable treatment on the four dependent variables, locus of control (LOC), the Modified Caution Index (MCI), the Person Average R (PAR), the Maximum Log Likelihood Function (lo), and a mathematics test (MT) are summarized in Table 11. Overall, the treatment was a factor which account for a statistically significant amount of variance in each of the four dependent variables. The kind of treatment a group received played an important part in determining the average magnitude of person reliability or response pattern indices of a group.

Table 11

Summary of the Analyses of Variance
for the Effect of the Treatment on Locus of Control (LOC),
the Modified Caution Index (MCI), the Person Average
R (PAR), the Maximum Log Likelihood Function (lo),
and a Mathematics Test (MT).

	LOC	MCI	PAR	lo	MT
F	13.20	20.19	18.64	11.61	1.31
P<	249E-8	34E-10	14E-9	114E-7	0.27

df=2, 573

Results of the Chi-Square Tests

The Chi-square test was performed on the data collected from the follow-up Survey questionnaire in order to examine the effect of the treatment on the subjects' responses to items associated with subject's locus of control orientation and test-taking behavior: subject's perception of the test difficulty level, the expectancy of the test outcome, the test strategy employed, and the attitude toward the test. All four Chi-square tests achieved statistically significant results which means that the kind of treatment or test instruction a group received played an important part in determining the frequency in each category. A detailed description of the results are presented below.

The result of the Chi-square test for the overall difference between the groups classified by the variables treatment and test difficulty level perception (Table 12) was statistically significant (X^2=67.757, df=2, p<.01). Subjects' perception of the test difficulty level was related to the kind of the treatment they received. Specifically, in the external group with negative test instruction there were more subjects (149, 77%) who considered the test to be difficult; while in the internal group with positive test instruction, considerably fewer subjects (71, 37%) considered the test difficult. In the control group more subjects than those in the internal group but less than those in the external group (122, 63%) thought the test was difficult. In other words, variable treatment made a difference in subjects' perception of the test difficulty level.

Table 12

The Chi-Square Test for the Effect of the Treatment on the Test Difficulty Level Perception

Treatment	Difficult	Not Difficult
External	149	43
Internal	71	121
Control	122	70
χ^2=67.757 df=2 p<.01		

The results of the Chi-square test for the overall effect of the treatment upon the expectancy of the test outcome are reported in Table 13. There was a statistically significant difference in the frequency distribution of each cell (χ^2=19.817, df=2, p<.01). Over half of the subjects in the external group (113, 59%) deemed that the outcome of the test was dependent on their ability, while in the internal group more subjects (150, 78%) felt that their abilities determined the outcome of the test. To say it differently, more subjects in the external group who received external instruction thought that the outcome of the test depended on their luck. Again, the variable treatment made a difference in subjects' expectancy of the test outcome.

Table 13

The Chi-Square Test for the Effect of the Treatment on the Test Outcome Expectancy

Treatment	Ability	Luck
External	113	79
Internal	150	42
Control	144	48
χ^2=19.817 df=2 p<.01		

The results of the Chi-square test for the overall effect of the treatment upon the use of test strategy, guessing are presented in Table 14. Again the frequency distribution in the six cells depended on the treatment, test instruction, the subjects had received (χ^2=20.28, df=2, p<.01). There were more subjects in the external group (180, 94%) admitting the use of guessing strategy than those in the control group (158, 82%) and in the internal group (149, 77%). What kind of test instruction, positive or negative, they received were reflected in the number of subjects who acknowledged the use of guessing strategy in test-taking.

Table 14

The Chi-Square Test for the Effect of the Treatment on the Use of Guessing Test Strategy

	Guessing	
Treatment	Guessing	No Guessing
External	180	12
Internal	149	43
Control	158	34

χ^2=20.28, df=2, p<.01

Table 15 presents the results of Chi-square test for the overall effect of the treatment upon the subject's attitudes towards the test. The treatment had no significant effect on the frequency distribution in each of the two categories of test attitude (χ^2=1.704, df=2, p>.05). In other words, the subject's attitudes towards the test were not influenced by the treatment he/she received. Among the three treatment groups the number of subjects who said that they were serious about the mathematics test was very close, the external group: 119, 62%; the internal group: 123, 64%; and the control group: 131, 68%.

Table 15

The Chi-Square Test for the Effect of the Treatment on the Attitude Toward the Test

Treatment	Serious	Not Serious
External	119	73
Internal	123	69
Control	131	61

$\chi^2=1.704$, df=2, $p<.01$

In summary, there are more students in the external group who considered the test difficult, their luck would determine the test outcome, and guessing was part of their test strategy than those in the control and internal groups, conversely, there are more students in the internal group who deemed that the mathematics test was not so difficult, their ability would determine the test results, and that guessing was not their test strategy. In other words, the kind of treatment or test instruction a group received could largely explain the difference between the groups.

CHAPTER 4

DISCUSSION

This study had two major purposes. The first was to determine the relationships between three response pattern indices (the modified caution index, MCI; the Person Average R, PAR; the Maximum Log Likelihood Function, lo) and two other variables: locus of control (LOC), and score on a mathematics test (MT). The second purpose of this study was to determine the effect of the contingency test instructions on subjects' locus of control orientation and their response pattern. The instructions were internal, external, and neutral intended to have positive, negative, and neutral effect respectively on subjects' locus of control orientation. It was expected that a shift in LOC would influence person reliability or response pattern.

The Relationships Between the Modified Caution Index (MCI), the Person Average R (PAR), and the Maximum Log Likelihood Function (lo)

The findings of the correlation study verified the first major hypothesis regarding the relationships between the three person reliability or response pattern indices, the Modified Caution Index (MCI), the Person Average R (PAR), and the Maximum Log Likelihood Function (lo). As expected, this study reported a strong negative correlation between MCI and PAR, and a moderate negative correlation between MCI and lo, and a moderate positive correlation between PAR and lo. The similarity of the underlying principles of the three response pattern indices explains the relationships between them, their magnitudes and their directions.

The Modified Caution Index (MCI) describes how different a subject's response pattern is from the response pattern norm derived from the whole group. In contrast, PAR examines the similarity between an individual's response pattern and that of his/her peers. IRT theory-based lo measures the degree to which each person's response pattern fits the model used to characterize normal behavior, or the contribution of a person's response pattern to the likelihood function.

Clearly these three response pattern indices have one thing in common: they all compare an individual's test-taking behavior with certain criteria, a norm, or a model, or his/her peers. The difference between them is which side of the picture of test-taking behavior is of interest to a particular index, difference (MCI) or similarity (PAR, and lo). From this point of view, it is not unexpected to find a high to moderate, negative and positive, correlations between MCI, PAR, and lo.

A closer examination of the underlying principles of the three indices would show that MCI and PAR appear to be more similar than they are with lo. For the sake of differentiation, MCI is labeled as group dependent and PAR classic theory based. To some extent, PAR can also be classified as a group dependent index. Instead of deriving a group norm first and then comparing (MCI method), PAR compares an individual's response pattern through Pearson's correlation procedure with his/her peer's and then computes the average, an approach different in order but similar in logic to the one adopted by MCI. A close relationship or a high correlation between MCI and PAR is thus expected. The same logic might be used to explain why MCI's and PAR's relationships with lo (lo is a somewhat sample independent index) are not as strong as the one between MCI and PAR. The relationships of MCI with lo and PAR with lo are about half of the magnitude of the correlation between MCI and PAR. It implies that the more abstract sample independent lo which focuses on the contribution of a person's response pattern to the maximum likelihood function features something different from what is a more direct measure of similarity between an individual's response pattern and that of the group measured by MCI and PAR.

A high correlation between MCI and PAR leads one to expect that both of them will have a similar degree of correlation with other variables, e.g., locus of control of this study. Conversely, it is expected that lo will differ from MCI and PAR in its relationships.

The Relationships Between Response Pattern Indices (the Modified Caution Index, MCI; the Person Average R, PAR; the Maximum Log Likelihood Function, lo) and Locus of Control (LOC)

An important finding of this study is that response pattern and locus of control are related. As hypothesized, LOC was positively related to MCI, and negatively related to PAR. LOC was negatively related to lo to a much less degree. The results indicate that a person with a more external locus of control orientation tends to have an unusual or aberrant response pattern. A person with a more internal locus of control orientation tends to have a normal or non-aberrant response pattern. The relationship between locus of control and response pattern is more evident when response pattern is measured by MCI and PAR than by lo.

Originally, the correlations between locus of control orientation and person reliability or response pattern indices were expected to be between 0.30 to 0.50 as most of the studies relating locus of control and other behaviors have reported. It was not expected that the obtained correlations between the locus of control measure (LOC) and the person reliability indices (MCI and PAR) would reach the 0.60's which means that PAR and MCI share approximately 36% of the variance with LOC. The immediate explanation seems to be that person reliability or response pattern is indeed related to locus of control.

In addition to the existence of a relationship between locus of control and person reliability or response pattern, a few other factors might have contributed to the finding of this higher than expected correlation obtained from this study. First, the mathematics test was moderately difficult for the target subjects as originally intended. It had a mean mathematics test score of 25 out of 40 items and a test difficulty level of .625. The test scores were widely distributed within a range from 10 to 37 points. One might summarize that difficult test items may have forced subjects to make decisions based more on their locus of control orientations, that is, internal students relying more on their efforts and external students more on their luck, thus polarizing their response patterns and increasing the range of response pattern indices of the whole study.

Second, LOC was widely distributed with a range from 2 to 20. The above range was larger than the LOC range reported in other studies possibly because of the joint effect of test instructions and a difficult test on their locus of control perception at the time immediately after taking the test. Large distributions of MCI, PAR, and LOC allow for maximum correlations between them.

Researchers have reported that both locus of control (LOC) and person reliability or response pattern indices (MCI, PAR, and lo) are related to achievement scores to various degrees (Coleman, Campbell, Hobson, Mcpartland, Mood, Weinfeld, & York, 1960; Harnish & Linn, 1981; Ayabe & Heim, 1987; Levine & Drasgow, 1982; and Shishido, 1991). Similar results were found in this study and reported in Chapter III. It is important to know what part the achievement score plays in determining the relationship between locus of control and response pattern. Partial and semi-partial correlation analysis indicated that the relationship between MCI and LOC and the relationship between PAR and LOC remained relatively unchanged when taking into consideration the fact that the score on a mathematics test (MT) correlated moderately with MCI, PAR, and LOC.

First, the above results lend further support to the finding from the overall group data that there truly exists a relationship between LOC and response pattern, when represented by MCI and PAR. Second and equally important, the relationship between locus of control and response pattern is not affected by persons' test scores. The correlation between locus of control and response pattern is not due to their respective correlations with the corresponding test score. Persons' score on a test, whether high or low, will not substantially alter the degree of the relationship between persons' response pattern indices (MCI and PAR) and LOC. There seems to be an unique overlap between response pattern indices, MCI and PAR, and LOC. This connection or overlap is not shared by the variable of test score and thus will not change with changes in test scores.

The finding that response pattern is independent of test scores sends a strong message that the information derived from response pattern will be unique, something not provided by the total test score. Total score can not differentiate subjects at the same score level. In this situation, response pattern and locus of control orientation are able to further differentiate subjects. Another implication of the finding is that in predicting an individual's locus of control, response pattern will produce a more accurate estimate than MT. Likewise, locus of control can predict response pattern better than MT.

As predicted in the discussion of the relationships between the three response pattern indices, unlike MCI and PAR, the partial and semi-partial correlations between lo and LOC dropped markedly from the original Pearson product-moment correlation which was modest at most, and neither of them was statistically significant. The results imply the score on a mathematics test (MT) plays a significant part in the degree of the relationship between lo and LOC. The overlap between lo and LOC is mostly shared by MT. Compared with the correlation between LOC and lo, the correlation between LOC and MT is relatively large. A relatively large correlation between LOC and MT and a weak correlation between lo and LOC result in a non-significant, close to zero correlation between lo and LOC. In other words, there is no independent connection or unique overlap shared by lo and LOC. lo will not be able to provide information about an individual's locus of control orientation beyond the information provided by test scores.

The overall group data includes two treatment groups, internal and external, it is possible that the results from the previous correlation analysis might be contaminated by the test instructions and that the relationship between response pattern indices and locus of control orientation might differ if the two treatment groups were excluded. As a precaution, the simple Pearson product-moment correlation, partial correlation, and semi-partial correlation procedures were repeated with the data from the control group which were not affected by external and internal test instructions. The results are similar to those obtained from the overall database. It appears that the moderate relationship of locus of control with response pattern, when measured by MCI and PAR, were not substantially affected by the application of the treatment, internal or external test instructions.

At this point one might conclude the following: first, a person's response pattern measured by MCI and PAR appears to be moderately related to a person's general locus of control orientation. Second, the moderate correlation between response pattern (when measured by MCI and PAR) and LOC is independent of test score and test instruction. lo has a modest correlation with locus of control.

Studies concerning response pattern in the past have focused mainly on the administrative or external factors, such as syllabus and instructional difference. Instead, this study focuses more on subjects' dispositional characteristic, locus of control orientation. Researchers and practioners can utilize the information derived from students' response pattern to predict their locus of control orientation in decision-making process. Further, if researchers can determine the direction of the relationship between them, they can train students to have more consistent response pattern, if desired, by developing programs to modify students' locus of control orientation.

The Effect of the Treatment on Locus of Control (LOC), and Three response pattern indices, the Modified Caution Indices (MCI), The Person Average R (PAR), and the Maximum Log Likelihood Function (lo)

The results of ANOVA largely confirmed the third and fourth hypotheses concerning the effect of the treatment on subjects' locus of control and response pattern indices, MCI, PAR, and lo. In general, the test instruction is a significant factor in determining the variate value. The differences between all possible pairs of group means were not all statistically significant, i. e., the effect of the test instruction were not equally reflected by the different indices.

The Effect of the Treatment on Locus of Control (LOC)

In the case of locus of control, external instruction succeeded in swaying subjects' locus of control orientation toward the external side of the continuum. It appears that internal test instruction also had some effect on subjects' locus of control orientation in the direction of internality.

The above finding can be explained, at least in part, by subjects' responses to the survey items. They were constructed such that subjects' response to the survey items (Table 16) will bring to light the ANOVA findings. They will help to explain why and how test instructions influence subjects' test perception, outcome expectancy, and test performance.

Table 16

Summary of the Frequency Distribution for the Effect of the Treatment on the Perception of Test Difficulty, Test Outcome, the Amount of Guessing, and Test Attitude

	Treatment		
	External	Control	Internal
Difficult	149	122	71
Not Difficult	43	70	121
Luck	79	48	42
Ability	113	144	156
Guessing	180	158	149
No Guessing	12	34	43
Serious	132	123	119
Not Serious	61	69	73

The Chi-square tests indicated that subjects' perception of test difficulty level, test outcome expectancy, and the use of guessing strategy were all related to the kind of test instructions they received. Compared with those in the control and internal group, more subjects in the external group agreed that the test was difficult, believed that their luck would determine the test outcome, and reported the use of guessing strategy in taking the test. One might argue that if a person thinks a test is difficult, that person is more likely to believe the outcome of the test will depend on his/her luck. It appears then, if a person takes a gambling attitude toward a test, that person tends to take chances and resort to guessing as a natural choice of test-taking strategy.

It seems, however that the messages conveyed by the test instructions affected the subjects differently. On the first survey item, both instructions were effective. Surprisingly, internal test instruction demonstrated a stronger effect on subjects' test difficulty level perception than the external test instruction.

The treatment effect on the test outcome expectancy differs from the treatment effort pattern on test difficulty level perception, more like the one on locus of control. The effect size of the external instruction for test outcome expectancy, 31 (the difference between the frequency of the external group, 79 and the frequency of the control group, 48) was slightly larger than the corresponding effect size for test difficulty level perception, 27 (the difference between the frequency of the external group, 149 and the frequency of the control group, 122). The effect size of the internal test instruction for test outcome expectancy 6 (the difference between the frequency of the internal group, 42, and the frequency of the control group, 48) was much smaller than the corresponding effect size for test difficulty level 51 (the difference between the frequency of the internal group, 71, and the frequency of the control group, 122). The reason might be that the expectancy of the test result was more difficult to change than the perception of the test difficulty level. The difference between the two survey items is that the perception of test difficulty level was more test content specific, and the test outcome expectancy was a more generalized concept. Researchers in the field of locus of control have reported that the higher order or more general a concept is, the harder it may be to change. This explanation may also apply to the test instruction effect on locus of control.

Another point worth mentioning is that the effect size of the external test instruction, 31 (the frequency difference between those who believed in luck in the external group and those in the control group) is much larger than the effect size of the internal test instruction, 6 (the frequency difference between those in the control group and those in the internal group). This might be explained such that under the assumption of equal treatment effect, subjects tended to accept more willingly the external test instructions which offered them an excuse for not working hard. It seems to be more difficult to urge people to work harder, or they do not usually work hard, or they do not have the ability even if they desire to do so. This may also explain the finding from ANOVA on the LOC data, i.e., it is easier to influence an individual's locus of control in the direction of externality than internality.

The overall effect of the test instruction for the use of guessing strategy in test-taking was similar to that on the test outcome expectancy. The effect size of the external test instruction, 22 (the difference between the frequency of the external group and the control group), continued to be larger than the effect size of the internal test instruction, 9 (the difference between the frequency of the internal group and that of the control group), more than doubling the size. The external test instruction succeeded in persuading more subjects to report the use of guessing test strategy while the internal test instruction succeeded, to some extent, in discouraging them to do so as expected. It sounds logical to explain that subjects' belief that chance plays an important role in the test led them to believe that what they needed to do was to make a guess and take a choice at random. An external instruction really made things easier for those who did not want to work too hard by providing them a good reason. The results from this item foretell there will be a good chance of obtaining some differences of MCI, PAR, and lo between the treatment groups, external and internal, and the control group.

In sum, the test instructions were created to be parallel in format. The results show that test instructions have different effects on LOC. It seems to be easier to make subjects more externally oriented than internally in this case. It might be interpreted such that subjects appeared to accept more readily the message conveyed by the external test instruction. The external test instruction emphasized that the test was so difficult that no matter how hard one would work it would not make much difference. Since the test instruction implies that the test outcome depended on one's luck, one might generalize his/her conclusion from the external test instruction to other events, after experiencing one hour and a half of long frustration of taking a difficult test. People tend to find some excuses other than their lack of ability and effort, for their average and/or below average performance so that they would at least feel better with the consequence. It is more so when such excuses were readily provided with no need for further justification. Working on a hard test was definitely not great fun for most of the young Chinese students who had been overwhelmed by a heavy study burden in preparing for the college entrance examination two years ahead. A lot of students would easily give up pursuing academic excellence if not for the mounting pressure from society, school, family, and parents in particular. On the other hand, under the same circumstance, subjects did not feel substantially more confident in their control of an event than the control group subjects just because of a piece of encouraging instruction. Students in China are under constant pressure and do not have much control of their study. Almost everything in school is programed for them. All they have to do is to complete the school schedule. Researchers have reported that people would have a higher sense of self-control only after achieving success especially through their own effort. One student said after taking the test, "How I could feel more confident in controlling an event when I was having a hard time with the test just a few minutes ago?"

Another possible explanation related to the above discussion is that external test instruction appears to be more effective in affecting subjects' locus of control orientation. Though efforts were made to write the two test instructions with equal effectiveness, i.e., parallel in wording, results have indicated that they did not work as expected. It is quite clear that a similar format alone would not produce a similar effect. The external test instruction fulfilled its mission better than the internal test instruction in most cases.

One area which needs careful study is the instrument employed in measuring the concept of locus of control. Rotter's Internal-External Control Scale seems to be an adequate instrument for measuring Chinese senior middle school students. One original concern of this study was whether Rotter's scale developed in the United States could measure Chinese students in a way as it was intended.

The difference between the two cultures and traditions, Chinese and American, are well known and well documented (Hsieh, Shybut, & Lotsof, 1969; Lai, 1977; Hui, 1982). A brief discussion might help to illustrate some points. Different cultures provide different kinds of rewards for social behavior. These systems of rewards influence an individual's attribution of causality toward the outcome of social events. An effective and appropriate response to social events in one culture may be entirely inappropriate in another. The contingency of reinforcement also may vary accordingly.

Western culture places high value on achievement through one's effects. Life experiences are supposed to be largely a consequence of one's action. In contrast, the traditional Chinese culture stresses kinship, the past, and the maintenance of the status quo. The individual is encouraged to view his life as being relatively fixed.

Understandably, individuals raised in a culture that value individualism, and personal output of energy are likely to be more internally oriented than individuals from a culture that tends to emphasize a different set of values. A study by Hsieh (1969) reported that subjects in the United States which emphasized independence tend to be more internally oriented than the American Chinese and Hong Kong Chinese subjects. It is expected that Chinese students tend to be more externally oriented according to the definition of locus of control developed by Rotter. The average score of the Chinese students on Rotter's scale was higher than the sample means of some large scale studies. In this sense, Rotter's scale appears to measure the general concept of locus of control consistently.

The translation of the instrument from English into Chinese might have played a part in ensuring its quality. A significant amount of time and energy was devoted to meeting this challenge. Three bilingual consultants worked on creating a Chinese version of the scale as accurate as possible.

As much as possible, they made some necessary modifications without sacrificing the original meaning in order to make the phrases understandable and the items answerable for the Chinese students who lacked knowledge of western culture.

In addition, the results of the fourth survey item indicate that in all three groups about two thirds of the students reported that they were serious about the test. It implies that the remaining one third of the students were not. If they took the whole test lightly, paid no or little attention to the instruction, and went straight to the test items, then the key of the study test instruction would have been useless. Since one third of the students did not cooperate, their response to the items might produce some ambiguity, compromise the test instruction effect, and cause difficulty in interpreting the results. Perhaps, performing analysis on data from only those "serious" students will produce cleaner results.

The Effect of the Treatment on the Person Reliability Indices: the Modified Caution Index (MCI) and the Person Average R (PAR), the Maximum Log Likelihood Function (lo), and a Mathematics Test (MT)

The results of the one-way ANOVA on response pattern indices confirmed the hypotheses that the treatment, in general, has an effect on subjects' person reliability or response pattern indices, MCI, PAR, and lo. As expected, the treatment was able to influence subjects' test-taking behaviors leading them in the direction intended by the test instructions respectively. However, the effectiveness of a specific test instruction on the indices varied as in the case of LOC. A summary of the Scheffé tests for the effect of test instruction on the indices is presented in Table 17.

The Scheffé test on the lo data yielded the same results as on LOC. The lo mean of the external group (-15.13) was significantly smaller than the lo mean of the control group (-14.31). The lo mean of the control group did not statistically differ from the lo mean of the internal group (-11.95). The implication seems to be that lo was able to reflect effectively the impact of the external test instruction, but failed to detect the effect of internal test instruction. At this point one may state that lo is sensitive to external test instruction, and can differentiate students with external instruction from those with control and internal instructions.

Recall that LOC and lo has a modest relationship. MCI has a much stronger relationship with LOC than lo. One would expect the effect of test instructions on MCI would be similar to that on LOC, but stronger than that on lo. However, tests report the opposite results, the MCI mean of the

Table 17

Summary of the Scheffé Test for the Effect of the Treatment on Locus of Control (LOC), Modified Caution Index (MCI), Person Average R (PAR), the Maximum Log Likelihood Function (lo), and a Mathematics Test (MT)

Treatment	LOC Grp Mean	MCI Grp Mean	PAR Grp Mean	lo Grp Mean	MT Grp Mean
External	A 12.43	A .23	A .22	A -15.13	A 25.50
Control	B 11.30	A .21	B .23	B -14.31	A 26.07
Internal	B 10.89	B .18	C .26	B -13.95	A 26.19

Means with the same letter are not statistically significantly different at the .05 level.

internal group (.18) was significantly lower than that of the control group (.21), while the MCI mean of the external group (.23) was not significantly higher than that of the control group. MCI reflected only the effect of the internal test instruction.

MCI and PAR have a very close relationship between them; it is expected that the effect of test instruction on PAR should follow the pattern of MCI. As expected, the PAR means of the three groups are all different statistically. Not only the difference between the external and control group, but also the difference between the control and internal group, are statistically different. PAR is a powerful index sensitive to both instructions in this study.

From the statistical point of view, several factors contribute to the power of a test. They are sample size, level of significance, treatment effectiveness, within-group variability, and instrument sensitivity. In the present study, subjects and sample size are held constant, 192 for each group, same subject pool for all measures. The test instruction treatment is also a constant. However, the measures of the treatment effect vary, each with its unique characteristic concentrating on certain areas in one general direction. Different measures with varying degrees of sensitivity to the same treatment result in different effect sizes. Besides, the measuring instruments employed affect the within-group variability. A sensitive measure will produce a larger difference between two group means and a smaller standard error which means a better chance of rejecting the null hypothesis. In other words, the response pattern index used may have played a role in determining the power of ANOVA in this study. Evidently, all three response pattern indices changed in the direction intended by the test instructions. Among them, only PAR reflected the effect of all the treatments with two balanced group differences and a small standard deviation.

The mathematics test score (MT) was the only variable which failed to reflect the full effect size of the treatment. The MT means of the three groups are different, but the differences are small. It implies that the kind of test instruction received by the groups did not have any significant effect on the subjects' mathematics test scores, though it changed subjects' locus of control orientation and response pattern indices at different degrees. To change a person's response pattern indices or locus of control dramatically does not necessarily produce a dramatic change in test score. Missing or getting one or two easy or difficult items would shift a subject's position from one end of the response pattern indices' continuum to the other. It supports the previous argument that subjects at the same score level might have different locus of control orientation and response pattern. Total test score provides limited amount of information about subjects. If more information about a test taker is desired, response pattern may be an option to explore.

To summarize to this point, the external instruction had an impact on subjects' locus of control orientation swaying them in the direction of externality. Among the indices, PAR is a sensitive measure of the test instruction effect. MCI reflected only the effect of the internal test instruction, and lo external test instruction. Subjects' mathematics test scores were not affected by either test instructions.

Conclusions and Recommendations

The conclusions of the present study are basically consistent with the hypotheses based on the literature review.

First, the measure of locus of control is related to response pattern indices. A person with a strong belief of the decisive role of outside factors such as luck, fate, and powerful others tends to have an aberrant response pattern. The instructions had some effect on subjects' general locus of control orientation and their test-taking performance. Those receiving external instruction were more externally oriented; the effect of the internal instruction on locus of control appeared to be weak. The effect of the test instruction on response pattern largely depends on the sensitivity of the individual index. Among the three employed in this study, PAR appears to be sensitive to both of the test instructions, while MCI and lo reflect only one of the two instructions, internal and external respectively.

Locus of control and response pattern are related and the instrument intended primarily to change subjects' expectancy altered not only subjects' locus of control orientation, but also their response pattern. There appears to exist a directional relationship between LOC and response pattern, i.e., subjects receive messages from test instruction, process it, modify their locus of control orientation, and act on it accordingly with response pattern as an end product in the test-taking situation.

The above conclusion is tentative, preliminary and by no means conclusive. The results must be interpreted with caution, and more studies need to be conducted to ascertain the findings.

The data of the study were collected after the treatment and the mathematics test were administered. The control group may not really be "treatment" free, for the control subjects received a "neutral" instruction which might thus change the situation they were in. A neutral test instruction could possibly have some effect on subjects' locus of control orientation and response pattern. Further study might be conducted to find the relationship between locus of control and response pattern without applying the treatment to subjects.

Phares (1957) reported that the treatment and task had a joint effect on subjects' expectancy. Researchers might examine the interaction effect of locus of control orientation and test instruction on subjects' response pattern. Any knowledge from it will improve the effectiveness of a training program aiming to modify a person's response pattern, the design of test instructions, test administration, and even teaching strategy.

The findings of the present study are based on the data collected from the Chinese senior middle school students in Shanghai with some unique characteristics. It would definitely be meaningful to have studies similar to this one replicated in countries such as the United States.

APPENDIX A.1

TEST INSTRUCTIONS

In English

A. Internal Instruction

We have given this test to a large number of students just like you. We know that this is a difficult test but within tenth grade capabilities. You might find many of the questions are answerable. This means, if you think things through and try very hard, you will be able to answer many of the questions correctly. Although we are quite sure that people are able to answer many of the questions correctly, we need a few more people to prove it statistically.

Important! Answer all questions.

B. External Instruction

We have given this test to a large number of students just like you. We know that this is a difficult test beyond the tenth grade capacities. You might find many of the questions are too difficult. This means, if you think things through and try very hard, you will still be unable to answer many of the questions correctly. Although we are quite sure that people are unable to answer many of the questions correctly, we need a few more people to prove it statistically.

Important! Answer all questions.

C. Control Instruction

We have given this test to a large number of students just like you. We know that this is a test of average difficulty level. However, some students have done very well, answering many of the questions correctly, while some students have done rather poorly, missing many of the questions. Since we want to make sure how difficult this test is, we need a few more students to prove it statistically.

Important! Answer all questions.

APPENDIX A.2

TEST INSTRUCTIONS

In Chinese

这份试卷我们已经给大量和你一样的学生做过。我们知道这份试卷很难，其难度超过高一学生的能力范围。你会发现许多题目太难了些。也就是说，即使你认真思考，努力答题，你还是不能够答对其中的许多题目。当然，我们比较确信学生不可能做对其中的大部题目，但是为了从统计角度上来证明这一点，所以我们现在还需要一些学生来试一下这份试卷。

注意！请回答所有题目。

这份试卷我们已经给大量的和你一样的学生做过。我们知道，这份试卷较难；但是，它的难度在高一学生的能力范围之内。你会发现许多题目都是可以做出来的。也就是说，如果你认真思考，努力解题的话，你能够答对其中大部分的题目。虽然我们比较确信学生们能正确地做出许多题目来，但是为了从统计角度上来证明这一点，我们还需要一些学生做一下这份试卷。

注意！请回答所有的题目。

这份试卷我们已经给大量和你一样的学生做过。我们知道这份试卷的难度属中等水平。但是，有些学生的答卷情况很好，做出了大部分试题；而有些学生的答卷情况很差，有许多题目做不出来。我们想弄清楚这份试卷究竟有多难，並从统计角度上来证明这一点，所以我们现在还需要一部分学生来试一下这份试卷。

注意！请回答所有题目。

APPENDIX B.1

A MATHEMATICS TEST

In English

PLEASE NOTE

Copyrighted materials in this document have not been filmed at the request of the author. They are available for consultation, however, in the author's university library.

127-135

137-146

148-151

153-157

159

161

University Microfilms International

APPENDIX B.2

A MATHEMATICS TEST

In Chinese

APPENDIX C.1

ROTTER'S INTERNAL-EXTERNAL LOCUS OF CONTROL SCALE

In English

APPENDIX C.2

ROTTER'S INTERNAL-EXTERNAL LOCUS OF CONTROL SCALE

In Chinese

APPENDIX D.1

Survey Questions

In English

APPENDIX D.2

Survey Questions

In Chinese

REFERENCES

Ayabe, H. I., & Heim, M. (1987). *How low can a KR20 reliability go?* Paper presented at the annual conference of the Hawaii Educational Research Association, Honolulu, Hawaii.

Blixt, S. L., & Shama, D. D. (1986). An empirical investigation of the standard error of measurement at different ability levels. *Educational and Psychological Measurement, 46*, 545-550.

Brislin, R. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & W. W. Lamber (Eds.), *Handbook of cross-cultural psychology*, (Vol. 2, pp. 389-445). Boston, MA: Allyn and Bacon.

Coleman, J. S., Campbell, E, Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. T. (1966). *Equality of educational opportunity*. Washington, DC: U.S. office of Education (Superintendent of Documents, Catalog No. FS5.238:38001)

Committee of AERA, APA, & NCME to Develop Standards. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Donlon, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group-determined item difficulties. *Educational Psychological Measurement, 42*, 105-113.

Drasgow, F. (1982). Choice of test model for appropriateness measurement. *Applied Psychological Measurement, 6*, 297-308.

Feldt, L. S., Steffen, M., & Gupta, N. C. (1985). A Comparison of five methods for estimating the standard error of measurement at specific score level. *Applied Psychological Measurement, 9*, 351-363.

Gulliksen, H. (1962). *Theory of mental tests*. (4th ed.). London: Wiley.

Guttman, L. (1941). The Quantification of a class of attributes: A theory and method of scale construction. In P. Horst, P. William, & L. Guttman (Eds.), *The prediction of personal adjustment*, (pp. 321-347). New York: Social Science Research Council, Committee on Social Adjustment.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Harnisch, D. L. (1983). Item response pattern: applications for educational practice. *Journal of Educational Measurement, 20*, 191-206.

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

Harrow, M., & Ferrante, A. (1969). Locus of control in psychiatric patients. *Journal of Consulting and Clinical Psychology, 33*, 582-589.

Hersch, P. D., & Scheible, K. E. (1967). On the reliability and validity of internal-external control as a personality dimension. *Journal of Consulting Psychology, 31*, 609-613.

Hoyt C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153-160.

Hsieh, T. T., Shybut, J., & Lotsof, E. J. (1969). Internal versus external control and ethnic group membership. *Journal of Consulting and Clinical Psychology, 33*, 122-124.

Hui, C. H. (1982). Locus of control: A review of cross-cultural research. *International Journal of Intercultural Relations, 6*, 301-323.

Jaeger, R. M. (1988). Use and effect of caution indices in detecting aberrant patterns of standard-setting judgments. *Applied Measurement in Education, 1*, 17-31.

James, W. H. (1957). *Internal versus external control of reinforcement as a basic variable in learning theory.* Unpublished doctoral dissertation, Ohio State University.

Julian, J. W., & Katz, S. B. (1968). Internal versus external control and the value of reinforcement. *Journal of Personality and Social Psychology, 8,* 89-94.

Julian, J. W., Lichtman, C. M., & Ryckman, D. B. (1968). Internal-external control and the need to control. *Journal of Social Psychology, 76,* 43-48.

Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement, 4,* 105-126.

Kilpatrick, D. G., Dubin, W. R., & Marcotte, D. B. (1974). Personality, stress of the medical education process, and changes in affective mood state. *Psychological Reports, 34,* 1215-1223.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2* 151-160.

Lai, C. T. C. (1977). *Cultural and control orientation a study of internal-external locus of control in Chinese and American-Chinese women.* Unpublished doctoral dissertation, University of California, Berkeley.

Lefcourt, H. M. (1966). Internal vs. external control of reinforcement: A review. *Psychological Bulletin, 65,* 206-220.

Lefcourt, H. M. (1976). *Locus of control: current trends in theory and research.* Hillsdale, NJ: Lawrence Erlbaum.

Lefcourt, H. M. (1984). *Research with the locus of control.* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum.

Levine, M. V., & Drasgow, F. (1982). Appropriateness measurement: Review, critique and validating studies. *British Journal of Mathematics and Statistics Psychology, 35,* 42-56.

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

Liverant, S., & Scodel, A. (1960). Internal and external control as determinants of decision making under conditions of risk. *Psychological Reports, 7,* 59-67.

Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement, 20*, 325-336.

Lord, F. M. (1965). A strong true score theory, with application. *Psychometrika, 30*, 239-270.

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

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

Mollenkopf, W. G. (1949). Variation of the standard error of measurement. *Psychometrika, 14*, 189-229.

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., pp. 80-203). Reading, MA: Addison-Wesley.

Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.). New York: McGraw-Hill.

Phares, E. J. (1957). Expectancy changes in skill and chance situations. *Journal of Abnormal and Social Psychology, 54*, 339-342.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*, (Whole No. 609).

Rotter, J. B., & Mulry, R. C. (1965). Internal versus external control of reinforcement and decision time. *Journal of Personality and Social Psychology, 2*, 598-604.

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

Ryckman, R. M., & Sherman, M. F. (1973). Relationship between self-esteem and internal-external control for man and woman. *Psychological Reports, 32*, 1106-1110.

Sato, T. (1975). *The construction and interpretation of S-P tables*. Tokyo: Meiji Tosho.

Samejima, F. (1981). Final report: Efficient methods of estimating the operating characteristics of item response pattern categories and challenge to a new model for the multiple-choice item. *Technical Report*. Knoxville, TN: Department of Psychology, University of Tennessee.

Shishido, J. (1991). *A study of selected response aberrance indices as alternatives to the standard error of measurement in interpreting individual test scores*. Unpublished doctoral dissertation, University of Hawaii at Manoa.

Tatsuoka, K. K., & Linn, R. L. (1983). Indices for detecting unusual response patterns: Links between two general approaches and potential applications. *Applied Psychological Measurement, 7*, 81-96.

Tatsuoka, K. K., & Tatsuoka, M. M. (1982). Detection of aberrant response patterns. *Journal of Educational Statistics, 19*, 215-231.

Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operations by the individual consistency index. *Journal of Educational Measurement, 20*, 221-230.

Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 560-620). Washington D.C.: American Council on Education.

van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology, 13*, 267-298.

Weiss, D. J. (1973). The stratified adaptive computerized ability test. *Research report* (73-3). Minneapolis: University of Minnesota, Department of Psychology. (NTIS No. AD-768 376)

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97-116.

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

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA.