Bhargava and Ishizuka's BI-Method: A Neglected Method for Variable Selection

SHING ON LEUNG
The University of Macau

JOHN SACHS
The University of Hong Kong

ABSTRACT. Quite often in data reduction, it is more meaningful and economical to select a subset of variables instead of reducing the dimensionality of the variable space with principal components analysis. The authors present a neglected method for variable selection called the BI-method (R. P. Bhargava & T. Ishizuka, 1981). It is a direct, simple method that uses the same criterion—trace information—used in ordinary regression analysis. The authors begin by discussing the nature and properties of the BI-method and then show how it is different from other existing variable selection methods. Because the BI-method originally was applied to small datasets that had little or no relevance to psychology or education, the authors apply it to large datasets with relevance to the psychological and educational literature. Of particular interest was the application of the BI-method to select a subset of items from a large item pool. Two practical psychometric examples with 49 and 108 items, respectively, showed that item subsets selected with the BI-method reflected the underlying structure of the whole item pool and that the scales based on those item subsets showed good reliability and predictive validity. The appropriateness of this item selection method within the context of the domain-sampling model is discussed.

Key words: BI-method, item selection, principal components, principal variables, variable selection

A COMMON THEME IN DATA ANALYSIS is the need to reduce a large number of variables to a small number of dimensions that capture most of the information in the data, as is done in principal components analysis, or to select a subset of these variables that seems to capture the important aspects of all the data. Often, such data reduction methods can greatly simplify subsequent data analy-

Address correspondence to: Shing On Leung, Faculty of Education, The University of Macau, Taipa, Macau, China. E-mail: soleung@umac.mo

sis in multivariate analysis of variance, canonical correlation analysis, discriminant function analysis, or regression analysis. Data reduction also can provide for a more parsimonious and meaningful summary of the data that takes into account all the interrelationships among the variables. When multicollinearity is present, data reduction can remove these collinear dependencies.

For example, in regression analysis, principal component regression offers the advantage that the components retained are uncorrelated and, thus, simplifies the interpretation of the regression coefficients. But the principal components themselves may not be easy to interpret. Furthermore, the choice of which principal components to retain is not as simple as it might first appear. One could choose the first few components that account for a large portion (e.g., 60%) of the variance in the data. This has the added advantage that the variance of the estimated regression parameters for these components will be small. In fact, the variance will be proportional to $1/(n\lambda_i)$ (Dunteman, 1989), where n is the sample size, and λ_i is the ith eigenvalue that equals the variance of the ith principal component. Because the first few components account for most of the variance in the data, these components will, therefore, show the smallest variance for their estimated regression parameters. This strategy excludes from the regression analysis those components that account for a small amount of variance in the data. But as Jolliffe (1986) noted, "low variance for a component does not necessarily imply that the corresponding component is unimportant in the regression model" (p. 135). In fact, components that account for only a small amount of variance in the predictor set, X, may account of a large amount of variance in the criterion, Y. Consequently, it may not be possible to simultaneously retain those components that account for most of the variance in the predictor set, X, and delete those that account for the least amount of variance and still end up with a set of variables that optimally predicts the criterion, Y (Jolliffe).

Principal Variables

Because the determination of which principal components to retain can be problematic, and because the retained components themselves can be difficult to interpret, McCabe (1984, p. 138) applied "principal components optimality criteria" and identified four methods by which a subset of the original variables could be selected with minimal loss of information. All four methods involve either maximizing or minimizing some function of the eigenvalues for the retained and deleted variables such that the $k + 1$ or $k + 2$ variables retained would explain about the same amount of variance as k principal components. McCabe referred to these retained variables as *principal variables*, and they have the advantage that one ends up working with a familiar set of variables that have obvious substantive meaning. But because the maximization or minimization criteria used are different, the subset of variables retained by each method need not be the same, so one is still faced with the problem of which method to choose. McCabe offered some help here in concluding that his Method 2 "[which uses] the percentage of variation explained (i.e., the trace information) appears to be most suitable" (p. 144).

However, application of the trace information criterion was proposed earlier by Bhargava (1980) and Bhargava and Ishizuka (1981) and referred to by them as the *BI-method for variable selection.* Furthermore, Bhargava and Ishizuka provided a direct, exact, and efficient algorithm employing the sweep operator (Jennrich, 1977), whereas McCabe did not provide a direct, optimal, and efficient algorithm for this criterion. The lack of optimality in McCabe's method occurs because it does not always work on the trace information. Rather it uses the determinant as an initial step to screen subsets of variables and only then evaluates them in terms of the trace information, whereas the BI-method always works on the trace information.

Jolliffe's (1972, 1973, 1986) B2 and B4 variable selection methods use either the discard or retained principal components, respectively, to identify principal variables. However, because Jolliffe's B2 and B4 methods do not consider the correlation among all variables, they can only be treated as ad hoc procedures, whereas the BI-method offers an objective criterion for selecting principal variables that does consider the correlation among all variables.

Trace Information

Because our orientation is applied, we shall not go into all the statistical theory underlying the BI-method. We present only the basics here; for more technical details, the reader is referred to Bhargava and Ishizuka (1981).

The aim of the BI-method is to select variables that share substantial variance with the discarded variables so that little information is lost. It does this by using the trace information as the basic criterion for variable selection. Let matrix Σ denote a $(p + q) \times (p + q)$ population covariance (or correlation) matrix whose variables can be written as the partitioned vector $\mathbf{X}' = (\mathbf{X}'_1, \mathbf{X}'_2)$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are vectors with dimensions p and q, respectively. Then the population covariance (correlation) matrix, Σ , can be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \qquad (1)$$

where submatrices Σ_{11}, Σ_{12}, and Σ_{22} have dimensions $p \times p$, $p \times q$, and $q \times q$. The trace information (tr) in the covariance (correlation) matrix, $tr\Sigma$, measures the total variation in $\mathbf{X}$, and is considered as the total information. If $\mathbf{X}_1$ is selected from $\mathbf{X}$, then the residual covariance (correlation) matrix of $\mathbf{X}_2$ on $\mathbf{X}_1$ is $\Sigma_{2.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ (see Morrison, 1990, pp. 91–92). The jth diagonal element of $\Sigma_{2.1}$ contains the residual variance in the regression of the jth variable in $\mathbf{X}_2$ conditional on all variables in $\mathbf{X}_1$. Therefore, $tr\Sigma_{2.1}$ measures the total variation in $\mathbf{X}_2$ conditional on $\mathbf{X}_1$, and it represents the information lost, and $tr\Sigma - tr\Sigma_{2.1}$ represents the information retained in $\mathbf{X}_1$. The *normalized trace information* retained in $\mathbf{X}_1$ is denoted by

$$I_{\mathbf{X}_1'} = 1 - \frac{tr\Sigma_{2.1}}{tr\Sigma}. \tag{2}$$

The normalized trace information ranges from zero to one and is the proportion of information accounted for by $\mathbf{X}_1$ from $\mathbf{X}$. It is directly comparable to the usual information used in ordinary principal component analysis. In principal component analysis, if p' principal components of $\mathbf{X}$ are retained, then the trace information in these retained components can be given in terms of their eigenvalues: $\Sigma_{i=1}^{p'}\lambda_i$, where λ_i is the ith eigenvalue of matrix Σ. The normalized trace information for these retained components is the proportion of the total variance in $\mathbf{X}$ accounted for by the p' principal components. If p variables and p' principal components contain the same amount of trace information, and if p is slightly larger than p', the reduction in the number of variables given by the BI-method might be more economical and meaningful. More important, principal components analysis does not reduce the number of variables because it uses their linear combinations. However, the BI-method does reduce the number of variables to be used.

To simplify the problem, when we wish to select a single variable, say the ith variable, we do not know what i is unless we obtain the minimum of $tr\Sigma_{2.1}$. Then for the ith variable, X_i, we have $tr\Sigma_{2.1} = \Sigma\,\mathrm{Var}(X_j)(1 - r_{ij}^2)$, where $\mathrm{Var}(X_j)$ is the variance of the jth variable X_j, and r_{ij} is the correlation between the ith and jth variable, with summation over all j except i. If we maximize the normalized trace information, I, hence minimize $tr\Sigma_{2.1}$, then we are selecting X_i such that it has the highest correlation with those variables not selected, that is, the largest r_{ij}^2.

Because the BI-method selects variables that are highly correlated with those not selected, it gives maximum results with the original variables, and hence will have the highest predictive power of the original variables. Not only does this simple and direct variable selection procedure retain most of the information in the original set of variables, it also uses the same criterion as in ordinary principal component analysis and regression analysis. Thus, the BI-method is a nonarbitrary procedure for selecting a smaller set of variables for further study from a larger variable pool.

BI-Method

Implementation of the BI-Method

The BI-method as described here uses a *forward stepwise procedure*, and one of the main components for implementing the BI-method is the sweep operator

(Jennrich, 1977), which is an efficient algorithm for finding the inverse of a covariance (correlation) matrix, especially when the matrix is partitioned. With the sweep operator, the quantity $tr\Sigma_{2.1}$ can be easily calculated. Different selection of variables will give different values of this quantity. Now, the aim of the stepwise procedure is to find the best selection of variables that can minimize this quantity and hence maximize the normalized trace information. The stepwise procedure is described as follows:

Step 1: Selection of the first variable. To select the first variable, we simply calculate the quantity $tr\Sigma_{2.1}$ for each variable and then choose the one with the minimum value.

Step 2: Selection of the second variable. Given that the first variable is selected, the second variable is selected based also on the minimum value of $tr\Sigma_{2.1}$.

Step 3: Min-max procedure of variable exchange. It is possible that a variable may be withdrawn after it has been selected because other combinations of variables, or an exchange between selected and not selected, can be profitable. To achieve this, a min-max procedure is used. If the maximum value of those variables not selected is bigger than the minimum of those variables selected, an exchange can be better. Details can be found in Bhargava and Ishizuka (1981).

Steps 2 and 3 will be repeated for selecting the third, fourth, and all other variables until all are included. As a result, for a given number of variables selected, we will have a good selection of variables, together with the corresponding normalized trace information. Although the normalized trace information increases with the number of variables selected, in most cases the marginal increase usually diminishes as the number of variables selected increases.

Readers interested in using the BI-method with their data can obtain an executable file from the authors on request. Further details on the algorithm can be found in Bhargava and Ishizuka (1981, pp. 36–38). Apart from the stepwise procedure for the BI-method just described, there is also an all-possible subset selection procedure developed by the authors, which will be published later (Sachs & Leung, 2006). However, the stepwise procedure is still the most useful when the number of variables (items) involved is large because the all-possible subset selection procedure is impossible in that situation. For example, with 10 variables, there will be 1,023 ($=2^{10} - 1$) distinct subsets. For 20, 30, 40, and 50 variables, this number will be roughly 10^6, 10^9, 10^{12}, and 10^{15}, respectively. In our example with the Harman (1976) dataset, we are working with 24 variables in 1 hr using a CPU of 1,395MHz. However, corresponding times for 30, 40, and 50 variables will be 3 days, 7 years, and forever! This dramatic increase in time occurs because, for each added variable, the number of possible subsets will be doubled. Consequently, the stepwise procedure is still practically the most useful and, therefore, is the one presented here.

Application of the BI-Method

Although the BI-method originally was applied to only small datasets that had no direct application to educational and psychological literature, we show its applicability to education and psychology using three large datasets from the psychological and educational literature that involve many variables. Because the scales of measurement of the variables in these examples are essentially arbitrary, we analyze the correlation matrices. The BI-method, however, can be applied to either the covariance or correlation matrix, but the variables selected from each matrix may not be the same, although there can be considerable overlap if the variables are measured on similar scales. In the three examples that follow, the standard BI-method used was the forward stepwise procedure as described by Bhargava and Ishizuka (1981).

The first dataset is a correlation matrix of 24 psychological tests (i.e., composite or scale-level measures) reported in Harman (1976, p. 124); the second dataset is the 50-item Coopersmith Self-Esteem Inventory (CSEI; Coopersmith 1967, 1981). The third dataset is the Fennema–Sherman Mathematics Attitude Scales with 108 items. The item-level datasets were chosen to see whether the BI-method could select a subset of items with acceptable reliability and good predictive utility (see McDonald, 1999, pp. 199, 222–223).

Results

Harman Data

The correlation matrix of the 24 psychological tests reported in Harman (1976) is based on a sample of 145 primary school children. All 24 tests were tests of mental ability that Harman (p. 125) classified into the following five groups: G1 = spatial relations (1, 2, 3, 4); G2 = verbal ability (5, 6, 7, 8, 9); G3 = perceptual speed (10, 11, 12, 13); G4 = recognition (14, 15, 16, 17, 18, 19); and G5 = associative memory (20, 21, 22, 23, 24). However, the factor results reported in Harman indicate that the G5 factor was poorly defined and was eliminated.

Principal component analysis on the correlation matrix of these 24 psychological tests extracted 5 components with eigenvalues greater than 1, which together accounted for just over 60% of the total variance. Ten components accounted for about 77.5% of the variance. The BI-method results for 10 selected tests are shown in Table 1.

Five tests selected with the BI-method accounted for just under 50% of the variance or about 14% less variance than did the principal component solution for 5 components. No G1 or G5 tests were included among the 5 tests selected with the BI-method. When 10 tests were selected, the BI-method accounted for 67.32% of the total variance, which is about 10.2% less than accounted for by the

TABLE 1. Tests Selected and Percentage of Variance Explained With the BI-Method for the Harman (1976) Data

Step	Test selected	Variance explained (%)
1	9	19.10
2	9 13	29.55
3	9 13 18	36.46
4	9 13 18 10	42.06
5	9 13 18 10 16	47.41
6	9 13 18 10 16 3	52.11
7	9 2 18 10 16 3 4	56.18
8	9 2 18 10 16 3 4 15	59.97
9	9 2 19 10 16 3 4 15 13	63.91
10	9 2 19 10 16 3 4 15 13 20	67.32

first 10 principal components. However, principal component analysis has to use all the tests (i.e., it uses linear combinations for all the tests). The 10 tests selected with the BI-method included at least 1 test from each of the 5 test groups. The correlation of these 10 selected tests with the 14 unselected tests was .87. Thus, the 10 tests selected with the BI-method do a good job of reflecting the structure in the data as originally specified by Harman and in retaining a good portion of the information in the data.

Chinese Version of the Coopersmith Self-Esteem Inventory (CSEI)

To demonstrate the practical uses of the BI-method in the area of psychometrics for selecting a subset of items from a larger well-defined item pool, we chose the CSEI as an example. The original inventory was developed by Coopersmith (1967, 1981), and a modified version translated into Chinese was developed by Young (1995) in Taiwan and was eventually adapted in Macau, China (Chan, 2002). The CSEI measures self-esteem in adolescent populations. It is a long inventory comprising 58 items, 8 of which form a Lie Scale. The other 50 items measure various aspects of adolescent's self-esteem using a dichotomous response format (*like me* or *not like me*). Although a total inventory score based on all 50 items of the CSEI provides a global measure of self-esteem, separate scores can also be obtained on three subscales: (a) social self–peers, (b) home–parents, and (c) school–academic (Chiu, 1988).

If a smaller number of items from the CSEI could be used to assess global self-esteem, this would have obvious practical advantages provided that the selected items have good predictive utility and acceptable reliability (typically, a reliability of at least .70 is recommended for research purposes; see Nunnally, 1978, p. 245). This is not an easy task to accomplish because selecting items to maximize the estimate of internal consistency reliability will mean that these items are highly correlated, and selecting items to maximize predictive utility means that they will have zero or at least low correlations with each other (McDonald, 1999, pp. 224–225). One way of achieving this seemly paradoxical objective of good predictive utility and acceptable reliability is to use inventory subscale scores that show good internal consistency but low between-subscale correlations (McDonald, p. 225). However, this does not offer a solution for reducing an item pool such that good predictive utility is obtained without undue sacrifice of reliability. The BI-method is useful here, provided one is dealing with a well-defined item-content domain.

The data consisted of the responses of 583 junior secondary students in Macau, to (a) the Chinese version of the CSEI (only 49 items were used because 1 item was found not to be applicable to the Taiwan and Macau context and was dropped), (b) two scales measuring their relationship with their mother and father (adapted from Wu, 1999), and (c) a final scale measuring life adaptation. The BI-method was used to select two smaller subsets of items from the CSEI: one containing 12 items and the other containing 25 items. The results are shown in Table 2.

First, we note that the percentage of information retained in the shortened scale, as measured by the trace information, is more than a quarter and more than a half for the quarter- and the half-length scales, respectively. Thus, shortening by a given proportion does not translate into loosing the same proportion of information.

TABLE 2. Results From the BI-Method Applied to the Coopersmith Self-Esteem Inventory (1981; N = 583)

	Scale length		
Measure	Quarter	Half	Full
Number of items	12	25	49
Trace information (%)	38.9	64.5	100.0
Internal consistency reliability (α)	.76	.82	.87
Scale correlations with			
Relationship with mother	.32	.36	.38
Relationship with father	.29	.32	.35
Life adaptation	.51	.57	.61

Note. One item was not applicable to the Macau context and was dropped. All correlations were significant at the .001 level.

Second, we note that the reliability of approximately .87 falls at the bottom end of the range for this inventory reported by Chiu (1988) and is almost identical to that reported by Herz and Gullone (1999). Third, we see that the reliability of the selected subset with 25 items is only slightly less than that of the original 49-item version of the inventory and that the other subset with 12 items still has acceptable reliability of approximately .76. Thus, using the BI-method to shorten scales does not seriously sacrifice reliability.

Finally, we note that those correlations produced by quarter- and half-length scales are very near to those based on the full-length scale. Furthermore, all correlations with external criteria were highly significant ($p < .001$). Thus, we conclude that a subset of items selected with the BI-method will have acceptable reliability and almost equivalent predictive utility as a larger subset of items. This is not surprising and can be explained by examining Table 3.

Inspection of the correlations among the quarter-, half-, and full-length scales in Table 3 shows them all to be very high. Hence, it makes only a little difference which reduced scale one uses when obtaining correlations with other traits because the BI-method selects items that will result in the highest correlation with the full-length scale, thereby ensuring high predictive utility. Furthermore, the BI-method can retain representations in subscales of the original scales, as reported in the Table 4.

The original CSEI consists of four overlapping subscales. As shown in Table 4, the number of items as selected with the BI-method is distributed quite evenly within each subscale. In the shortened form of the scale, each subscale is well represented. Hence, the BI-method can select items that reflect the subdomains in the original full scale.

Fennema–Sherman Mathematics Attitude Scales

To demonstrate the stability of the BI-method, we applied it to another large dataset. The sample consists of 538 students in Grades 7 and 9 in Macau. The

TABLE 3. Correlations Among Quarter-, Half-, and Full-Length Scales of the Coopersmith Self-Esteem Inventory (1981)

	Scale length		
	Quarter	Half	Full
Quarter		.93	.88
Half	.93		.93
Full	.88	.93	

Fennema–Sherman (1976) Mathematics Attitude Scales was applied to measure the students' attitude toward mathematics. The scale comprises nine domains or subscales: (a) usefulness of mathematics; (b) attitude toward success in mathematics; (c) confidence in learning mathematics; (d) effectance motivation in learning mathematics; (e) mathematics as a male domain; (f) mathematic anxiety; and (g) father's, (h) mother's, and (i) teacher's attitude toward learning mathematics. Results very similar to those found with the CSEI are reported in Table 5.

Based on the full scale of 108 items, three shortened scales were obtained with the BI-method—sixth, quarter, and half scales—in which the number of items selected are in proportion accordingly. The amount of information retained (i.e., trace information) by the shortest scale with only one sixth of the items (only 18 items) was nearly half (47.2%) of the total information, whereas the percentage of information retained for the quarter- and half-length scales was 56.7% and 77.9%, respectively.

The estimate of internal consistency reliability, by its nature, decreases as the number of items decreases. However, even the shortest scale with only 18 items had an acceptable reliability of .76. The corresponding reliability estimates for quarter- and half-length scales were .81 and .89, respectively. These are relatively high for attitude scales.

Inspection of the correlations between the shortened scales and the full-length scale shown in Table 6 reveals little difference between these correlations; they were all above .94. This suggests that shortened scales whose items are selected with the BI-method should correlate about as highly with an external criterion as the full-length scale would.

TABLE 4. Number of Items in Each Subscales for Quarter-, Half-, and Full-Length Scales of the Coopersmith Self-Esteem Inventory (1981) Selected by the BI-Method

	Scale length		
Subscales (overlapping)	Quarter	Half	Full
Self-rejection/self-affirmation	4	9	19
Parental approval	5	7	10
Rejection by authority	3	8	21
Social and self-acceptance	3	7	15

Note. The Coopersmith Self-Esteem Inventory has overlapping subscales so that some items may belong to more than one subscale. Therefore, summing the number of items for each subscale will not match the grand total for the scale length specified.

TABLE 5. Numbers of Items, Percentage of Trace Information, Reliabilities, and Correlations for Subscales for Sixth-, Quarter-, Half-, and Full-Length Scales of the Fennema–Sherman Mathematics Attitude Scales (1976) as Selected by the BI-Method

	Scale length			
Measure	Sixth	Quarter	Half	Full
Number of items	18	27	54	108
Trace information (%)	47.2	56.7	77.9	100.0
Reliability	.76	.81	.89	.95
Correlation with subscales				
Usefulness of mathematics	.61	.57	.65	.65
Attitude toward success in mathematics	.43	.48	.49	.45
Confidence in learning mathematics	.64	.61	.69	.71
Effectance motivation	.65	.63	.70	.72
Mathematics as a male domain	.45	.51	.52	.47
Mathematics anxiety	.62	.59	.61	.66
Father's attitude	.75	.77	.72	.74
Mother's attitude	.75	.78	.73	.76
Teacher's attitude	.66	.70	.66	.70

TABLE 6. Correlations Among Sixth, Quarter-, Half-, and Full-Length Scales of the Fennema–Sherman Mathematics Attitude Scales (1976)

	Scale length			
	Sixth	Quarter	Half	Full
Sixth		.961	.942	.946
Quarter	.961		.961	.954
Half	.942	.961		.981
Full	.946	.954	.981	

Table 7 reports the length of each shortened scale along with the distributions of the items selected from each of the item subscales. We can see that the subscales are quite evenly represented within each of the shortened scales. Even the shortest scale with only 18 items includes at least 1 item from each subscale. Hence, the BI-method appears to select items so that all the subdomains from the original full scale are represented.

TABLE 7. Numbers of Items in Each Subscale for Sixth-, Quarter-, Half-, and Full-Length Scales of the Fennema–Sherman Mathematics Attitude Scales (1976) as Selected by the BI-Method

	Scale length			
Subscale	Sixth	Quarter	Half	Full
Usefulness of mathematics	2	2	7	12
Attitude towards success in mathematics	2	4	7	12
Confidence in learning mathematics	2	2	6	12
Effectance motivation	1	2	5	12
Mathematics as a male domain	2	4	8	12
Mathematics anxiety	2	3	5	12
Father's attitude	3	4	7	12
Mother's attitude	3	4	5	12
Teacher's attitude	1	2	4	12
Total	18	27	54	108

Discussion

Existing methods of variable selection are provided by Jolliffe (1986) and McCabe (1984). Jolliffe's method depends on a somewhat arbitrary ad hoc criteria. McCabe provides four criteria for variables selection. Although his second criterion is basically the same as the BI-method, his algorithm begins by using his first criterion to screen variables and only then applies his second criterion. Hence, his algorithm does not employ an optimizing criterion.

In contrast, the algorithm for the BI-method uses the trace information from the beginning as its optimizing criterion. Furthermore, this is the same criterion used in ordinary regression and principal component analysis. Therefore, the BI-method has the virtue of being simple and direct, and it provides a unique solution for selecting principal variables.

However, there are two general criticisms that can be made of all such variable selection methods. The first is that these methods are blind to variable content because they use an optimization procedure that is seen as capitalizing on chance aspects of the sample data, thus questioning the validity of sample generalizations. A second criticism has to do with the characteristics of the sampled item subset. Here the objection is that the selected item subset may no longer represent the trait being measured by the whole item pool. That is, the selected item subset is seen as somehow changing the trait being measured.

The first criticism is not unique to variable selection methods and can be directed at other multivariate procedures that use optimizing criteria, such as principal component analysis, stepwise regression, and optimal scaling (i.e., dual scaling; Nishisato, 1994), to name but a few. Furthermore, even the most common univariate statistics (means, standard deviations, correlations, regression coefficients, etc.) are optimal sample quantities because they employ a least-squares property (Nishisato, 1979). Therefore, concerns about the validity of sample generalizations can be leveled at almost all statistical procedures if samples are poorly chosen.

The second criticism that the trait being measured may somehow be changed when a subset of items is selected from a larger item pool does not agree with the domain-sampling model (Guttman, 1945, 1953; Kaiser & Michael, 1975; McDonald, 1978, 1985, 1999, 2003; Nunnally, 1978; Tryon, 1957). In this model, the item content or behavior domain is seen as comprising a universal set of all items of a given kind (i.e., what the researcher is interested in measuring, whether self-efficacy or attitude toward mathematics, as in our examples). Typically, the universe of content (item-content domain) is conceptualized as countably infinite and so referred to as a *hypothetical universe of content* from which items can be drawn at random. Guttman (1944) noted that the content universe consists of all items that define the concept one is interested in (i.e., the common content) so that an item belongs to this universe by virtue of its content. Furthermore, Kaiser and Michael showed that no limiting assumptions about the internal structure of the items in the content domain need be made. That is, the content domain may or may not be unidimensional (Nunnally). More important, there is nothing in the domain-sampling model to suggest that composites constructed from randomly sampled item subsets would measure a trait different from that defined by the behavior domain from which the items were sampled. In fact, the whole purpose of the domain-sampling model is that the sampled item subsets reflect the infinite universal item-content domain (Nunnally; McDonald, 2003), and Guttman (1953) suggested that as few as 10 to 15 items should provide a close approximation to it.

With our data, we demonstrated that the shortened scales obtained with the BI-method were highly correlated with the original full-length scale, as seen in Table 3 for the CSEI data and in Table 6 for Fennema–Sherman Mathematics Attitude Scales data (1976). These results suggest that the trait measured by the item subsets is not very different from the trait measured by the full-length item scale.

In general, then, for a scale to have good psychometric properties, it has to be reliable and valid. In other words, its items should provide a good representation (sample) of the behavior or content domain (McDonald, 2003). Quite often, however, it may not be an easy job to attain both targets of high reliability and validity, and hence, some compromise has to be made. But once the validity and reliability of a scale are established, this is no longer an issue, and the task shifts to the selection of a subset of items that retain most of the information in the scale. In this article, we have demonstrated that the BI-method tends to select items from a large item pool that reflect the underlying structure of the item pool and that such selected item subsets provide good predictive validity and acceptable reliability.

In summary, the BI-method provides an objective way for selecting principal variables that retain most of the information in the data. Our examples show that the BI-method works well in a number of practical situations. Like regression analysis, the BI-method also uses the sweep operator. Hence, the existing step-wise procedure can be generalized to all possible subset variables selection.

REFERENCES

Bhargava, R. P. (1980). *Selection of a subset of pollution stations in the Bay Area of California on the basis of the characteristic, 24-hour suspended particulate concentration, from the viewpoint of variation* (Tech. Rep. No. 37, 1–22). Under the auspices of the Siam Institute for Mathematics and Society. Department of Statistics, Stanford University, Stanford, CA.

Bhargava, R. P., & Ishizuka, T. (1981). Selection of a subset of variables from the viewpoint of variation—An alternative to principal component analysis. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions* (pp. 33–44). Calcutta, India: Indian Statistical Institute.

Chan, I. W. (2002). *A study of self esteem and its corresponding factors for junior secondary school students in Macau.* Unpublished master's thesis (in Chinese), University of Macau, China.

Chiu, L. (1988). Measures of self-esteem for school-age children. *Journal of Counseling and Development, 66*, 298–301.

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco: Freeman.

Coopersmith, S. (1981). *Coopersmith Self-Esteem Inventory (School Form).* Palo Alto, CA: Consulting Psychologists Press.

Dunteman, G. H. (1989). *Principal components analysis.* Newbury Park, CA: Sage.

Fennema, E., & Sherman, J. A. (1976). Fennema–Sherman Mathematics Attitude Scales: Instruments designed to measure attitudes towards the learning of mathematics by males and females. *JSAS Catalog of Selected Documents in Psychology, 6*(1), 3b.

Guttman, L. (1944). A basis for scaling quantitative variables. *American Sociological Review, 9*(2), 130–150.

Guttman, L. (1945). A basis for analyzing test–retest reliability. *Psychometrika, 10*, 255–282.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika, 18*, 277–296.

Harman, H. H. (1976). *Modern factor analysis* (rev. 3rd ed.) Chicago: The University of Chicago Press.

Herz, L., & Gullone, E. (1999). The relationship between self-esteem and parenting style: A cross-cultural comparison of Australian and Vietnamese Australian adolescents. *Journal of Cross-Cultural Psychology, 30*(6), 742–761.

Jennrich, I. R. (1977). Step-wise regression. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Statistical methods for digital computers: Mathematical methods for digital computers* (Vol. 3, pp. 58–75). New York: Wiley.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. I: Artificial data. *Applied Statistics, 21*, 160–173.

Jolliffe, I. T. (1973). Discarding variables in a principal component analysis. II: Real data. *Applied Statistics, 22*, 21–31.

Jolliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Kaiser, H. F., & Michael, W. B. (1975). Domain validity and generalizability. *Educational and Psychological Measurment, 35*, 31–35.

McCabe, G. P. (1984). Principal variables. *Technometrics, 26*, 137–144.

McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability." *Educational and Psychological Measurement, 38*, 75–79.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillside, NJ: Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McDonald, R. P. (2003). Behavior domains in theory and in practice. *The Alberta Journal of Educational Research*, *49*(3), 212–230.

Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

Nishisato, S. (1979). Dual scaling and its variants. In R. E. Traub (Ed.), *Analysis of test data (New Directions for Testing and Measurement quarterly sourcebook;* pp. 1–12). San Francisco: Jossey-Bass.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis*. Hillsdale, NJ: Erlbaum.

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

Sachs, J., & Leung, S. O. (2006). *Constructing shortened versions of Fennema–Sherman Mathematics Attitude Scales employing a trace information criterion.* Manuscript submitted for publication.

Tryon, R. C. (1957). Reliability and behavior domain validity reformulation and historical critique. *Psychological Bulletin*, *54*(3), 229–249.

Wu, Y.-S. (1999). *The verification of the self-esteem model of adolescents*. Unpublished master's thesis (in Chinese), National Taiwan Normal University, Taiwan.

Young, S.-P. (1995). *A study of relationship among attachment relationships: Self-esteem and career development of adolescence*. Unpublished master's thesis (in Chinese), National Taiwan Normal University, Taiwan.