# Shortened Versions of Fennema-Sherman Mathematic Attitude Scales Employing Trace Information

John Sachs

Hong Kong

and

Shing On Leung

The University of Macau

Send correspondence to: Prof. Shing On LEUNG, J514, Silver Jubilee Building, Faculty of Education, The University of Macau, Taipa, Macau, China

Email: soleung@umac.mo

Abstract

This paper aimed at providing a simple and direct method for shortening the lengthy 108-item Fennema-Sherman Mathematics Attitude Scales (FSMAS). The method proposed uses the trace information (widely used in other popular procedures such as principal component analysis and regression analysis) as the criterion for item selection. Results for half-length FSMAS versions (54 items) constructed using the trace-information criterion compared favourably with Muhern and Rae's (1998) 51-item FSMAS short form. But unlike the later short form, our method retained as much variance as possible in the original 108-items and maximized the correlations, and hence predictive validity, with the full-length FSMAS version. The appropriateness of selecting items from a larger item pool based on employing a trace-information criterion is discussed within the context of the domain-sampling model.

*Keywords:* BI-method; Domain sampling; Item selection; Stepwise; Trace-information; item selection

## Introduction

For more than two decades, one of the most popular instruments for assessing attitudes toward the learning of mathematics by females and males has been the Fennema-Sherman Mathematics Attitude Scales (FSMAS) (Fennema & Sherman, 1976a,b) (Hyde, Fennema, Ryan, Frost, & Hopp, 1990). The FSMAS consists of nine domain-specific, 5-point Likert-type scales, that measure: (1) Attitude toward the Success in Mathematics Scale (S), (2) Mathematics as a Male Domain Scale (MD), (3) and (4) Mother (M)/Father (F) Scale, (5) Teacher Scale (T), (6) Confidence in Learning Mathematics Scale (C), (7) Mathematics Anxiety Scale (A), (8) Effectance Motivation Scale in Mathematics (EM), and (9) Mathematics Usefulness Scale (U) (Fennema & Sherman, 1976b).

FSMAS factor analysis results reported by Fennema and Sherman (1976a) (excluding the Anxiety Scale because of its high correlation with the Confidence Scale) identified four factors for males and females. A subsequent factor analytic study of the FSMAS by Broadbooks, Elmore, Pedersen, and Bleyer (1981) on the eight scales (excluding Anxiety) for females and males both individually and together yielded only one factor, while a similar analysis by them of all nine scales yielded only two factors. However, their item-level analysis of the FSMAS did identify eight factors that basically supported the theoretical structure of domain-specific scales as described by Fennema and Sherman (1976a). A further item-level factor-analytic study by Melancon, Thompson, and Becnel (1994) using principal component analysis with varimax rotation also provided generally favorable results supporting the validity of the scores from the FSMAS. However, a similar study conducted by Mulhern and Rae (1998), identified only six factors suggesting that the FSMAS might not be measuring the domain-specific scales exactly as specified by Fennema and Sherman (1976a,b). Mulhern and Rae (1998) went on to develop a shortened version of the FSMAS (FSMAS-SF) that contained slightly less than half the items.

The need for a shortened version of the FSMAS is evident from the length of the original instrument which contains 108 items (12 items on each of nine scales) and takes about 45 minutes for a typical respondent to complete. Fatigue alone would call into question the validity of responses to items near the end of the instrument. Development of Mulhern and Rae's (1998) FSMAS short form (FSMAS-SF) involved two stages. In the first stage, factor analysis was used to establish the internal structure of the whole scale. In the second stage, items were selected based on the results of the first stage. This two stage process yielded a 51-item FSMAS-SF in which nine items with the highest factor loadings were retained from each of the first five factors extracted and six items retained with the highest factor loadings on the sixth factor. (A more detailed discussion of Mulhern and Rae's factor analysis method will be presented later.) While this two-stage factor analysis method can be used to construct short forms of the FSMAS, there is a more simple, direct and objective way to do obtain short forms of any instrument.

The method we propose is the neglected BI-method of Bhargava (1980) and Bhargava and Ishizuka (1981). Originally developed for applied science and technology applications, the BI-method selects items using the trace information criterion and can be applied to either the correlation or covariance matrix. Therefore, the purpose of our study was to apply Bhargava and Ishizuka's (1981) BI-method to develop short forms of the FSMAS that would show acceptable reliability and good predictive validity and then to compare the BI-results against Mulhern and Rae's (1998) factor analysis results. While the BI-method implements item selection using the familiar stepwise and forward selection procedures, we were also interested in extending it to employ an all-possible subset selection procedure, dubbed the All-pos method, and compare these results, too.

## Method

### *Trace Information*

The BI and All-pos methods presented here use the trace information as a criterion. Actually, it is not a new concept because it is the same criterion as used in many popular data reduction methods in social sciences and psychology, e.g., principal component analysis, multiple regression analysis, canonical correlation, etc. Most data reduction methods use this as a criterion because it represents total variation information, or total variance, of a set of items or variables. Technically, trace information is the sum of all diagonal elements of the covariance or correlation matrix.

The difference between a common data reduction method such as principal component analysis and BI method is that principal component analysis reduces the dimensionality of the data by retaining the first few components accounting for an acceptable proportion of the total variance, but the retained principal components are still linear combinations of ***all*** the items. In contrast, the BI method selects items directly. The difference between multiple regression and BI method is that the former has dependent variables as external criteria. The BI method, like principal component analysis, has no external criteria, but the original total scale is the target aimed at in item selection. This ensures that the BI-method selects items that will have maximum correlation with the original scale, and hence good predictive validity. Though principal component analysis, multiple regression analysis and BI method differ in terms of selecting items (variables), they still all use the basic trace information criterion.

Because the BI methods presented below use the trace information criterion directly, they do not depend on any secondary results as is the case when employing Mulhern and Rae's (1998) factor analysis method. Details on this point are presented in the discussion section.

### *The BI-method and All-pos method*

The aim of the BI-method is to select items that share substantial variance with the discarded items so that little information is lost. It does this by using the trace information of the correlation (or covariance) matrix as the basic criterion for item selection. Here, trace information refers to the ratio of the information provided by the retained items over the total information, and, therefore, can range from zero to one. The BI method uses a forward stepwise selection process which involves three major steps.

Step 1. The first item is selected by calculating trace information associated with each item and choosing the one with the largest value.

Step 2. The second item is selected in the same way from remaining the items of step 1.

Step 3. Since a forward stepwise selection process is employed, items selected at an earlier stage are constantly compared against non-selected items so that an exchange will occur between selected and non-selected items in later steps if it is profitable. That is, it is possible for an item to be withdrawn after it has been selected and replaced by an earlier non-selected item.

Technically, the third step involves a "min-max" procedure for item exchange. If the maximum value of the trace information of those items not being selected is bigger than the minimum of those selected, the procedure will execute an exchange.

Steps 2 and 3 are iterated until all items are included based on the trace information criterion. Thus, for a given number of items, there will be good selection of items together with the corresponding trace information. While the information increases with the number of items selected, in most cases the marginal increase usually diminishes as the number of items increases. Details on the algorithm can be found in Bhargava and Ishizuka (pp. 36-38, 1981), and an executable file can be obtained from the authors upon request.

Though exchanges between selected and non-selected items are possible through out the item selection process, the standard BI-method is still a ***forward stepwise procedure*** in the sense that it selects new items based on what has been selected. However, there is also an all-possible subset selection procedure, the All-pos method, developed by the authors. But the All-pos method is very time consuming and calculates the trace information of ***all possible combinations*** of items one by one and then chooses the one with the largest value for a given number of items. While the All-pos method will give the best solution under the same trace information criterion, it is not practical when the number of items involved is large. This is because the All-pos method iterates over all possible $2^p$ sub-sets, where $p$ is the number of items, which makes it impossible if $p$ is large. For example, using a Pentium II home computer, 20 items have to go through $10^6$ iterations which takes five minutes, 30 and 40 items have to go through $10^9$ and $10^{12}$ iterations, respectively, and would take 4 days and 10 years! Even though we can increase the speed of the CPU, say by ten times, the geometric demand in time still cannot be satisfied. It is easy to see that for each additional item the number of possible patterns will be doubled. Hence, the BI-method's stepwise implementation is still the most practical when more than 30 items are involved.

In summary, the BI-method and All-pos methods are superior to other item selections methods because (a) they employ the widely used trace information criterion which means that the selected items show maximal correlation with the original scale (measure) and hence have maximal predictive validity, (b) they are simple and direct methods, and (c) they are objective since they do not rely on any arbitrary selection criteria such as factor loadings whose magnitude exceeds 0.4, say. Technical details on the BI methods can be found in Bhargava (1980), Leung and Sachs (2005), and Sachs and Leung (2005).

*Participants*

The sample consisted of 538 secondary one to three students (i.e., grades 7 to 9; age 13 to 15) from four typical secondary schools in Macau. One whole class was selected randomly from each grade (in grade 7, 8 and 9) in each school, resulting in 12 classes selected. The gender composition was 332 males (61.7%) and 191 females (35.5%), with 15 missing gender cases (2.8%).

*Instrumentation*

All participating children completed the 108 items of the nine Fennema and Sherman Mathematics Attitude Scales (FSMAS) during regular class hours. Additionally, demographic data was collected on children's gender and their family's socio-economic status in Macau.

*Analyses*

Both the BI-method and All-pos methods were used to select items for the FSMAS shortened versions. While the BI-method has no technical limitations with 108 items, the All-pos method as described above is practically impossible. Therefore, implementation of the All-pos method required that it be restricted to selecting items from within each of the nine scales comprising the FSMAS (12 items per scale) and then forming the overall FSMAS shortened versions by summing up all these shortened scales. Consequently, the All-pos method provided equal representation of each of the nine scales in the final shortened versions while the BI-method, which works on all 108 items of the full-length version at once, could not guarantee equal representation of each scale in the final FSMAS shortened versions.

*Comparison of factor analytic results between previous and present study*

To explore the characteristics of the current sample, we used the same factor analysis method employed by Mulhern and Rae (1998) to compare their results with those of the present study. Both studies identified six factors, and the percentages of variance accounted were 45.7% in Mulhern and Rae and 41.5% for the present study. Eigenvalues from correlation matrices ranged from 22.7 to 3.1 in Mulhern and Rae and 20.5 to 3.3 for the present study. Varimax rotation was used after factor extractions. Minimum factor saliency criterion of |.40| was used for factor interpretations, i.e., items were allocated to factors if factor loadings exceeded |.40|. This criterion resulted in 94 and 95 items out of the 108 items identified from

Mulhern and Rae and the present study, respectively. Mulhern and Rae reported five doublets (items identified with two factors) but the present study had none.

In Mulhern and Rae (1998), the first factor was identified with the Confidence (C) and Anxiety (A) scales and was interpreted as Mathematics-Related Affect scale. This was similar to the second factor of the present study, which included not only Confidence and Anxiety, but also Effectance Motivation (EM). While in Mulhern and Rae, there was no Effectance Motivation (EM) factor. Thus the Mathematics-Related Affect scale of the present study was more substantial represented than in Mulhern and Rae. The second factor in Mulhern and Rae was related to Father (F) and Mother's (M) Attitude and interpreted as Parent's Attitude. The first factor of the present study was similar to this factor but had an extra scale on Teacher Attitude (T), which was interpreted as Important Others' Attitude. In Mulhern and Rae, the Teachers' Attitude scale showed up as a separate sixth factor with fewer items. For both studies, the third and fourth factors were respectively Usefulness (U) and Male Domain (MD). The fifth factor in Mulhern and Rae was Success (S), but in the present study, it split into positive and negative scales on the fifth and sixth factors. But because the two-factor split of the Success scale in the current study occurred on the last two factors, their contribution to total variance accounted for was minimal. Interestingly, however, the factor analysis of the data in the present study was able to identify positive and negative items of the Success scale.

Whether a study reports a clear factor-structure pattern depends on the number of items with salient loadings on the corresponding factors allocated, and that should be as much as possible. In Mulhern and Rae (1998), the first factor had 92% (22 out of 24) of the factor loadings that exceed the |.40| criterion. The corresponding figures for the second to sixth factors were respectively 88% (21 out of 24), 75% (9 out of 12), 100% (12 out of 12), 75% (9 out of 12), and 50% (6 out of 12). Though the percentage of variance associated with the first factor was high, it was low for the last factor. The problem with the last factor was that only six items showed salient loadings. On the other hand, in the present study, figures for the number of salient loading exceeding |.40| were generally higher for the first to sixth factors: 81% (29 out of 36), 92% (33 out of 36), 100% (12 out of 12), 83% (10 out of 12), 100% (6 out of 6) and 83% (5 out of 6). For items with factor loadings exceeding |.40| but which did ***not*** correspond to the factors allocated, 12 such items were found in Mulhern and Rae but only two in the present study.

To conclude, the present study showed a clearer factor pattern than reported in Mulhern and Rae. This was probably due to a bigger sample size for the present study – 538 versus 196 reported in Mulhern and Rae (1998). Nevertheless, both studies produced generally similar factor structures for FSMAS, but the pattern was much clearer in the present study. We now report the results of the item selection methods based on the trace-information criterion and compare these results with those from the factor analysis.

## Results

### *Items Selected and Scale Content*

Using their factor analysis method to develop a shortened FSMAS, as previously noted, Mulhern and Rae (1998) selected nine items from each the first five factors based on items with factor loadings exceeding |.40|, while six items were selected from the last factor since only six items had loadings exceeding |.40|. So Mulhern and Rae's shortened FSMAS contained a total of 51 items. But, the representation of the original FSMAS on each factor in this shortened version was not the same as the original. The first two factors were represented by two scales each – C, A for the first factor and F, M for the second factor – while the other factors were represented by only one of the original scales each – U, MD, S and T, respectively for the third to sixth factors. Consequently, Mulhern and Rae's shortened FSMAS version did not have equal representation of the items from the original full-length FSMAS. This is a consequence of the arbitrariness associated with selecting items for inclusion based on Mulhern and Rae's factor analysis method, and will be discussed later. As Table 1 shows, (second last column) more items were selected from the Usefulness (U), Male Domain (MD) and Success (S) scales.

A similar outcome was observed when the factor analysis method was applied to the data in the present study. The first two factors were each represented by items from three of the original scales (F, M, T for the first factor and C, A, EM for the second factor). The third and fourth factors were represented by items from the U and MD scales, respectively, while the fifth factor was represented by S scale items with positive loading (+S) and the sixth factor represented by S scale items with negative loading (-S). In order to have as equal an item representation as possible from each scale, we would need to select items in proportion to number of scales allocated. Hence, the numbers of items selected in each factor would need to be in the sequence: 18, 18, 6, 6, 3 and 3, making up a total of 54 items for a half-length FSMAS version. But since within each factor items were selected according to the largest factor loadings (all exceeding |.40|), equal representation could not be secured – recall that items from three scales were allocated to the first two factors. Thus, selecting items according to the factor loading criterion cannot guarantee scales within the same factor to have equal number of items. This is reflected in the last column of Table 1.

As noted previously, implementation of the All-pos trace-information method meant that it could not be applied to the whole scale but rather had to be applied separately to the items within each of the nine scales. Thus by definition each shortened scale based on the All-pos method would have proportional item representation in all shortened FSMAS versions. That is, there would be two, three and six items in each scale for the one-sixth-, quarter- and half-length FSMAS versions. Hence, the content domain for each scale of the shortened FSMAS versions based on the All-pos method would remain the same proportionally to the scales in the original full-length FSMAS. By contrast, the BI-method would work on all 108 items at once, so proportional item representation for each shortened scale version would not be guaranteed.

Therefore, an obvious question is "If the BI-method is used to shorten the FSMAS, how would the item representation of its nine scales be affected?" Would it be significantly different from the proportional representation? Table 1 shows the results for the BI-method. As can be seen, the items were in fact quite evenly distributed among the nine scales for the one-sixth-, quarter- and half-length versions. Thus the number of items in each shortened scale did not appear to deviate too much from the expected values of two, three and six items per scale for the shortened one-sixth-, quarter- and half-length versions, respectively. The non-statistically significant chi-square values ($df = 8$, $p > .05$, in each case) of 2.00 (for 1/6 length scale), 2.67(for 1/4 length scale), and 2.33 (for 1/2 length scale) confirmed this. For half length scale, though non-significant values were observed from factor analysis (chi-square values of 10.6 and 5.5 respectively from factor analysis of previous and present studies), much smaller value of 2.33 from the BI-method showed that it gives a more even distribution. Hence, the BI-method was able to select items for shortened scales so that they had nearly proportional item representation of full-length version.

Table 1 also shows the number of items retained from each of the nine original scales based on the shortened 51-item version from factor analysis of previous study. The biggest difference between Mulhern and Rae's (1998) FSMAS-SF results and those for the BI-stepwise half-length version was for the Effectance Motivation (EM) scale where the BI method included five items, while Mulhern and Rae's study included only one item. With the exception of the Anxiety (A) scale where both shortened versions included five items, the differences between the numbers of items included in the other scales by Mulhern and Rae and the BI-method half-length version was plus or minus one to two items. For the scales based on factor analysis results in the present study, the distribution of items was quite uneven. The fourth and ninth scales had only two and four items respectively, and this resulted in lower predictive validity, as we shall see later. The exact items selected for each shortened scale can be provided upon request.

Insert Table 1 about here

*Percentage of Trace Information*

Table 2 shows the total number of items for each shortened version of the FSMAS obtained with either the standard BI-method or the All-pos method along with the corresponding percentage of trace information. By definition, the number of items is proportional to the percentage of items selected, but the percentage of trace information is usually larger than the proportion of items selected. For example, the half-length versions accounted for just over three-fourths of the total trace information while the one-sixth-length and quarter-length versions accounted for around 40% and 55%, respectively, of the total trace information.

Because the "All-pos" method as labeled in Table 2 does not correspond to all possible subset selection among 108 items all at once but rather to all possible subset selection from each of the nine 12-item scales comprising the FSMAS, each shortened version obtained with the All-pos method is comprised of the sum of the these nine All-pos shortened-scales. This explains why the percentage of trace information for the BI-method is larger than for the "All-pos" method. However, it is the same on half-length version, with the BI-method only slightly better for quarter-length versions, but the BI-method better for the shortest one-sixth-length version – 47% versus 38%.

Insert Table 2 about here

*Reliability Analysis*

In general, shortening a measure will lower its reliability. This is evident in Table 3 where the overall alpha reliability of the half-length version (0.90) is lower than for the full-length version (0.96). Nevertheless, this is still quite high though a little lower than the alpha of 0.93

based on factor analysis. A similar pattern is seen in the reliability estimates for the scale scores comprising the shortened version. While alpha reliability estimates for the half-length scale scores are all lower than those for the full-length scale scores, most still were greater than 0.70, with only two exceptions for Success (S) and Male Domain (MD) scales. It is noted that in Table 3 that Fennema and Sherman (1976) reported split-half reliability instead of coefficient alphas. With both half-length versions, the method of item selection based on factor analysis has higher reliability than the All-pos method. This is partly because for some shortened scales there were a larger number of items included. But more importantly, it is because with the factor analysis method, items are selected based on the factor loadings on each factor. Hence, those items with larger factor loadings within a factor are selected which results in greater internal consistency and, therefore, higher reliability. Thus methods of item selection based on a trace information criterion may result in shortened scales that show slightly lower internal consistency reliability. However, as we shall see in the next section, such shortened scales will show better predictive validity which out-weighs the disadvantages of slightly lower alphas.

Insert Table 3 about here

*Correlations and Predictive Validity*

An important property for any item-selection method based on trace information is that such shortened scales will have high correlations with the original scale. Table 4 shows the correlations of the FSMAS full-length version with the shortened one-sixth-, quarter- and half-length versions obtained with the BI-method and the All-pos method. As can be seen, scores for the shortened versions show nearly perfect correlations with scores on the full-length version no matter which method was used to select items. Even when we used only one-sixth (with 18 items) of the whole scale, the correlation with the original scale is as high as 0.95. Therefore, from the perspective of the domain-sampling model (Guttman, 1945, 1953; Tryon, 1957; Kaiser & Michael, 1975; Nunnally, 1978; McDonald, 1978, 1985, 1999, 2003), the sampled items in the shortened versions reflect the item-content domain as embodied by the full-length version. Consequently, the predictive validity for the whole scale for these shortened versions should be practically the same as for the full-length version.

Insert Table 4 about here

Correlations between the scale scores from the full- and quarter-length versions are presented in Table 5. Results for the one-sixth and quarter-length versions are not included because of few items. For full scales, i.e. the last row, both the All-pos method and factor analysis have high predictive validity (correlation of 0.97 ~ 0.98). For individual scales, correlations between the scale scores from the full- and quarter-length versions ranged from 0.86 to 0.90, while those from the full- and half-length versions ranged from 0.90 to 0.98. Thus there was good correspondence between scale scores of shortened and full-length version. If we compare the half-length version based on trace information (All-pos) with the factor analysis version (in the last column), we see that most correlations are as high as 0.9 with the original scale. The only exceptions occur for the Effectance Motivation and Teacher Attitude scales, where the results from the factor analysis method show notably lower correlations. This is not surprising because these two scales have the least number of items. The general pattern is that the All-pos method has higher predictive validity confirming that methods based on trace information can secure higher predictive validity.

Insert Table 5 about here

## Discussion and theoretical comparison

Well known methods for shortening scales are usually based on results from factor analysis, which is a popular method for analyzing the internal structure of scales. This kind of method involves two stages. The first is to fix the factor structure and select items based on salient factor loadings, or otherwise. The second is to decide on how many items to select from each factor. But these factor analytic methods have two types of arbitrariness associated with them. The first is to establish a clear factor structure for the scale that shows salient item-factor loading of each item usually exceeding |.40| on only one factor. But this is not always the case. The factor structure can vary from sample to sample with some items moving around on the factors or showing doublet loadings or some items may not showing salient loading on any factor. So, selecting items for shortened scales can be meaningful only if we can get a good, clear factor structure in the first instance.

The second kind of arbitrariness, even when the factor structure is clear, concerns the process of deciding how many items to select from each factor. In the present study, the number of items selected was done according to the number of scales allocated. This is reasonable but in some sense arbitrary, too. Even if this item selection method is accepted, the content balance for some scales may not be achieved if more than one scale is allocated to one factor. That explains why only two and four items were selected from the Effectance Motivation and Teacher Attitude scales with the factor analysis method reported in Table 1.

Apart from such arbitrariness, there is a difference in criterion for item selection. Methods based on factor analysis secure predictive validity by taking items from different factors, as different factors account for different proportion of variance, and hence trace information. But, within the same factor, it selects items according to the factor loadings. Selecting items based on factor loadings selects items that are more internally consistent ***with those selected***. This will increase the internal consistency and hence reliability of the selected items, but at the expense of predictive validity of the original scale. On the other hand, methods based on trace information are simple and direct and use trace information all the way through. Thus items are selected that are the most representative of the original scale, or more consistent ***with the original***, which maximizes the predictive validity. This was illustrated by results presented above, and also by the fact that only 16 out of 51 items overlap between All-pos method and method based on factor analysis in the present study.

Though shortened FSMAS versions obtained by selecting items with either the BI-method or the All-pos method have lower reliabilities than those based on factor analysis, they showed good properties in terms of the information retained and correlations with the original scale. Furthermore, the BI-method and All-pos methods can be implemented to obtain even shorter versions (one-third or one-quarter length). Comparatively speaking, construction of shortened scale versions based on factor analysis may not have this property because some factor loadings may not meet the minimum |.40| criterion for inclusion. In contrast the BI selection methods use a very clear and simple selection criterion – trace information – which ends up selecting those items that are most representative of the original scale in a non-arbitrary way.

With our data, we have demonstrated that the shortened scales, for any length, based on trace information were highly correlated with the original full-length scale as seen in Table 4. We can explain this as a consequence of maximizing the trace information, as occurs in ordinary regression analysis. These results suggest that the items included in our shortened versions do in fact reflect the same trait(s) measured by the FSMAS full-length version. The half-length versions of the FSMAS constructed with both the BI- and All-pos methods accounted for 78% of the variance of all 108 items and correlated 0.98 with the full-length FSMAS. Thus the BI-methods provided an objective way for selecting items which retained most of the information in the data and reflected the underlying structure of the FSMAS item pool quite well providing acceptable reliability and good predictive validity.

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. *Technical Report*, No. 37, *Siam Institute for Mathematics and Society*, 1-22.

Bhargava, R. P. and 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.

Broadbooks, W. J., Elmore, P. B., Pedersen, K., & Bleyer, D. R. 1981. A construct validation study of the Fennema-Sherman mathematics attitudes scale. *Educational and Psychological Measurement*, **41**, 551-557.

Fennema, E., & Sherman, J. A. 1976a. Fennema-Sherman Mathematics Attitudes Scales. *JSAS: Catalogue of Selected Documents in Psychology*, **6**, 31. (Ms. No. 1225).

Fennema, E., & Sherman, J. A. 1976b. Fennema-Sherman mathematics attitudes scales: instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal of Research in Mathematics Education*, **7**, 324-326.

Guttman, L. 1944. A basis for scaling quantitative variables. *American Sociological Review*, **9**, 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*, *IB*, 277-296.

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

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. 1990. Gender comparisons of mathematics attitudes and affect: a meta-analysis. *Psychology of Women Quarterly*, **14**, 299-324.

Jennrich, I. R. 1977. Step-wise regression. In K. Enslein, A. Ralston and H. S. Wilf (Eds.), *Statistical Methods for Digital Computers: Vol. III, Mathematical Methods for Digital Computers* (pp. 58-75). New York: John Wiley & Sons.

Leung, S. O., & Sachs, J. 2005. Bhargava and Ishizuka's BI-method: A Neglected Method for Variable Selection. *The Journal of Experimental Education*, **73**, 353-367.

Melancon, J. G., Thompson, B., & Becnel, S. 1994. Measurement integrity of scores from the Fennema-Sherman mathematics attitudes scales: the attitudes of public school teachers. *Educational and Psychological Measurement*, **54**, 187-192.

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: Lawrence Erlbaum.

McDonald, R. P. 1999. *Test theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum.

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

Mulhern, F., & Rae, G. 1998. Development of a shortened form of the Fennema-Sherman mathematics attitudes scales. *Educational and Psychological Measurement*, **85**, 295-306.

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

Sachs, J., & Leung, S. O. 2005. *A neglected method for variable selections*. Paper presented at the IMPS 2005 psychometric conference on July 5-8, 2005, Tilburg University, The Netherlands.

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

Table 1.

Number of items selected in each FSMAS scale from BI stepwise one-sixth, quarter and half length scale; and from Factor analysis in previous study (Mulhern and Rae 1998) and present study.

| Scales | BI One-sixth | BI Quarter | BI half | Factor analysis (previous study) | Factor analysis (present study) |
|---|---|---|---|---|---|
| Usefulness of Mathematics | 2 | 2 | 7 | 8 | 6 |
| Attitude Toward Success in Mathematics | 2 | 4 | 7 | 9 | 6 |
| Confidence in Learning Mathematics | 2 | 2 | 6 | 4 | 8 |
| Effectance Motivation | 1 | 2 | 5 | 1 | 2 |
| Mathematics as a Male Domain | 2 | 4 | 8 | 9 | 6 |
| Mathematics Anxiety | 2 | 3 | 5 | 5 | 8 |
| Father's Attitude | 3 | 4 | 7 | 6 | 7 |
| Mother's Attitude | 3 | 4 | 5 | 3 | 7 |
| Teacher's Attitude | 1 | 2 | 4 | 6 | 4 |
| Total | 18 | 27 | 54 | 51 | 54 |

Table 2.

Number of Items and Percentage of Trace Information for each Shortened Scale

| Scale Length | Number of Items | Percentage of Trace Information | |
|---|---|---|---|
| | | BI-method | All-pos Method |
| One-sixth | 18 | 47% | 38% |
| Quarter | 27 | 57% | 54% |
| Half | 54 | 78% | 78% |
| Full | 108 | 100% | 100% |

Table 3.

Coefficient alpha reliabilities of FSMAS scale for full- and half-length versions.

| Scale | (Fennema & Sherman, 1976a)(1) | Full-length (present study)(2) | All-pos method half-length (2) | Factor analysis half-length version (2) |
|---|---|---|---|---|
| Usefulness of Mathematics | 0.88 | 0.88 | 0.73 | 0.84 |
| Attitude Toward Success in Mathematics | 0.87 | 0.77 | 0.52 | 0.72 |
| Confidence in Learning Mathematics | 0.93 | 0.87 | 0.68 | 0.87 |
| Effectance Motivation | 0.87 | 0.88 | 0.71 | 0.73 |
| Mathematics as a Male Domain | 0.87 | 0.81 | 0.58 | 0.81 |
| Mathematics Anxiety | 0.89 | 0.88 | 0.72 | 0.88 |
| Father's Attitude | 0.91 | 0.88 | 0.72 | 0.87 |
| Mother's Attitude | 0.86 | 0.87 | 0.71 | 0.87 |
| Teacher's Attitude | 0.88 | 0.88 | 0.72 | 0.80 |
| Total scale | - | 0.96 | 0.90 | 0.93 |

(1) Note that Fennema and Sherman reported split-half reliability coefficients.
(2) These reliabilities are based on data of the previous study.

Table 4.

Correlations between Full FSMAS and Shortened Scales obtained with the BI-stepwise and All-pos Methods

| Scale Length | BI-method | All-pos Method |
| --- | --- | --- |
| One-sixth | 0.95 | 0.95 |
| Quarter | 0.95 | 0.97 |
| Half | 0.98 | 0.98 |
| Full | 1.00 | |

Table 5.

Correlations between each scale in the FSMAS full length version with (i) All-pos method quarter-length, (ii) All-pos method half-length, and (iii) factor analysis half-length version

| Scales | All-pos method quarter-length Version | All-pos method half-length version | Factor analysis half-length version |
|---|---|---|---|
| Usefulness of Mathematics (U) | 0.89 | 0.96 | 0.96 |
| Attitude Toward Success in Mathematics (S) | 0.86 | 0.93 | 0.93 |
| Confidence in Learning Mathematics (C) | 0.90 | 0.96 | 0.96 |
| Effectance Motivation (EM) | 0.88 | 0.94 | 0.84 |
| Mathematics as a Male Domain (MD) | 0.86 | 0.98 | 0.93 |
| Mathematics Anxiety (A) | 0.86 | 0.97 | 0.97 |
| Father's Attitude (F) | 0.90 | 0.95 | 0.95 |
| Mother's Attitude (M) | 0.89 | 0.95 | 0.95 |
| Teacher's Attitude (T) | 0.86 | 0.96 | 0.87 |
| Full Scale | 0.97 | 0.98 | 0.97 |