Journal of Educational Enquiry, Vol. 4, No. 2, 2003

Development and validation of a brief mathematics attitude scale for primary-aged students

Elaine Chapman
University of Western Australia
Crawley, Western Australia, Australia

Abstract

In this report I describe the development and validation of a 10-item scale for assessing primary-aged students' attitudes towards mathematics. A validation of the scale with 774 primary students indicated that the scale comprised three components: a reflection of students' overall enjoyment of mathematics; their perceptions of its value as a subject area; and perceptions of their ability to cope with their assigned mathematics work. The three components of the current scale were generally found to have acceptable reliability and validity across grades four, five and six. Recommendations for refinements to the scale and for its use with primary–aged students are discussed.

Introduction

Recent years have seen a revival of interest in the relationship between cognitive and affective factors in mathematics education (eg Fennema 1996). In particular, many reports have now been published on studies addressing the relationship between students' attitudes towards and achievement in mathematics, gender differences in mathematics attitudes, and the role that attitudes may play in explaining differences between boys and girls in their persistence with further study in the area (eg AAUW 1992; Frost et al 1994; Ma & Kishnor 1997). While the results of these studies have been somewhat mixed, several have indicated significant relationships between attitudes and achievement in math during the early and middle school years.

Given these findings, it is important that educators have access to instruments that are suitable for assessing mathematics attitudes in primary level students. In general, attitudes are defined as favourable or unfavourable dispositions toward social objects (Greenwald & Banaji 1995). More specifically, Neale (1969, p 632) defined mathematics attitudes as aggregated measures of 'liking or disliking of mathematics, a tendency to engage in or avoid mathematical activities, a belief that one is good or bad at mathematics, and a belief that mathematics is useful or useless'.

Of the various attitude scales that have been developed, the Fennema-Sherman Mathematics Attitudes Scales (FSMAS) (Fennema & Sherman 1976) remain the most extensively used in research studies (Hyde et al 1990). The FSMAS comprise nine scales of 12 items each (108 items in total). These assess attitudes towards Success in Mathematics, Mathematics as a Male Domain, Confidence in Learning Mathematics, Mathematics Anxiety, Effectance Motivation, and Usefulness of Mathematics. They also include Mother, Father and Teacher scales. The subscales can be used as a set, or individually. Recent studies on the psychometric properties of the FSMAS have generally provided support for the reliability and validity of the scales (eg Melancon et al 1994).

Aiken's Mathematics Attitudes Scales (MAS) (Aiken 1974; Aiken & Dreger 1961) have also enjoyed some popularity in mathematics education research. This instrument comprises two subscales of 10 items each, which assess students' enjoyment of mathematics (eg 'Mathematics is enjoyable and stimulating to me'), and their perceptions of its value as a subject area (eg 'Mathematics is a very worthwhile and necessary subject') – the E and V scales, respectively. Like the FSMAS, the technical adequacy of the MAS has been supported in a number of evaluation studies (eg Watson 1983). Taylor (1997) also argued that the relative simplicity and brevity of the MAS offered significant advantages to educators and to researchers in the field.

Much of the research that has used and/or evaluated these scales has, however, been focused on assessing attitudes at the university or high school levels. Due to the content and complexity of item wording in both instruments, neither is generally appropriate for use with younger students. For example, the 1974 version of the MAS includes questions such as 'Mathematics is not important for the advancement of civilization and society', while the FSMAS includes items such as 'Mathematics makes me uncomfortable, restless, irritable, and impatient'. Although the length of the two MAS and the individual FSMAS scales is not excessive, younger students can find it difficult to maintain their interest even with a 20-item scale. As noted by Mulhern & Rae (1998), maintaining students' interest levels is a significant problem when using longer attitude scales. Further, the complexity of the item wording in each instrument renders these scales unsuitable for use with students who have reading difficulties.

Adjective-based attitude measures are generally more economical than statement-based measures in terms of time requirements, and their use enables researchers to avoid some of the problems encountered in testing students who have reading difficulties. The generic School Subjects Attitudes Scales (SSAS) developed by Nyberg and Clarke (1979) present students with a list of 24 bipolar adjectives, each of which represent one of three attitude dimensions. The Evaluation scale reflects students' overall evaluative response to the subject area (eg nice–awful), while the Usefulness and Difficulty scales assess specific perceptions of the utility and difficulty level of the subject (eg important–unimportant and complicated–simple, respectively). An extensive evaluation across the fifth, eighth and eleventh grades supported the robustness of the factor structure and the reliability of the scales across several subject areas.

Despite these advantages, students can have difficulties in using adjective pairs to describe their reactions to academic subjects. For example, the SSAS includes adjective pairs such as ‘bright–dull’ (for the Evaluation factor), ‘smart–stupid’ (for the Usefulness factor), and ‘light–heavy’ (for the Difficulty factor). Responses to this scale thus rely heavily on students to make non-literal interpretations of the adjective pairs (eg students are expected to interpret ‘heavy’ as denoting ‘cognitively demanding’, rather than ‘weighty’). As such, the meaning of the scale items can be ambiguous. Thus, while adjective-based measures offer advantages in terms of reducing literacy demands and time requirements, statement-based measures offer advantages in terms of subject specificity and relevance.

The aim of this study was to develop a measure suitable for assessing students’ Mathematics attitudes in the primary grades. The goal in designing the 10-item ‘How I Feel About Maths Scale’ (HIFAMS) was to take advantage of the specificity of the statement-based approach, but to minimise the demands made on students’ literacy levels by simplifying the wording of statements. Students were asked to rate their agreement with each HIFAMS item on a five-point scale (from strongly agree to strongly disagree). The 10 scale items and their abbreviated labels are listed in Table I.

Table I. HIFAMS items and abbreviated labels

Item no.	Full item statement	Item label
1	Maths is boring.	Boring
2	Maths is too confusing.	Confusing
3	I enjoy my maths lessons.	Enjoy
5	Maths is an important subject.	Important
4	I can’t keep up with the work we do in maths.	Keepup
6	I like maths.	LikeMaths
7	I like maths more than my other school subjects.	Likemore
8	Doing maths problems is fun.	Probfun
9	I can’t see why I have to do maths.	Seewhy
10	Maths is a useless subject.	Useless

Participants

Participants in the validation sample were drawn from 14 schools that serviced a representative sample of areas in Western Australia. Based on data from the 1996 population census (Australian Bureau of Statistics 1997), three of these areas were allocated socioeconomic advantage indexes that fell within the upper 10th percentile; three within the lower 10th percentile; three within the lower 25th percentile; and five areas fell within the average range. Of this initial sample, six students (all from grade five) were excluded for failing to complete all 10 items of the scale. Mahalanobis distances (computed separately for each of the four grade levels) identified a further 13 score sets (three from grade four, four score sets from grade five, and six sets from grade six) as multivariate outliers ($\alpha = 0.001$), and these were also excluded from the final data pool. Table II shows median ages, grade levels, and numbers of males and females for the resulting sample of 774 students. Means and standard deviations for the 10 scale items are presented in Table III.

Table II. Grade, sex, and age data for final sample students

Grade level	Median age	Male N	Female N
4	9 years, 7 months	57	49
5	10 years, 3 months	142	152
6	11 years, 3 months	163	151
7	12 years, 9 months	18	42

Table III. Means (Ms) and standard deviations (SDs) for HIFAMS item scores

Item No.	Item	Label	M	SD
1	Maths is boring.	Boring	0.73	1.39
2	Maths is too confusing.	Confusing	0.73	1.25
3	I enjoy my maths lessons.	Enjoy	0.74	1.39
5	Maths is an important subject.	Important	1.46	0.99
4	I can't keep up with the work we do in maths.	Keepup	0.99	1.25
6	I like maths.	LikeMaths	0.45	1.43
7	I like maths more than my other school subjects.	Likemore	0.36	1.53
8	Doing maths problems is fun.	Probfun	0.30	1.28
9	I can't see why I have to do maths.	Seewhy	1.13	1.25
10	Maths is a useless subject.	Useless	1.34	1.15

Factor structure of the scale

To explore the factor structure of the scale, item responses for 793 students in grades 4–7 were intercorrelated and subjected to a principal components analysis (PCA). Correlations between the 10 HIFAMS items for the pooled sample are presented in Table IV. The PCA on this matrix indicated three components with eigenvalues greater than one, which together accounted for 62.9% of the total score variance. Initially, this solution was rotated to approximate a simple structure using both oblique (Direct Oblimin) and orthogonal (Varimax) procedures, with similar outcomes for the two methods.

Table IV. Intercorrelations between HIFAMS items

Item	1	2	3	4	5	6	7	8	9	10
Boring (1)	1.00									
Confusing (2)	0.36	1.00								
Enjoy (3)	0.31	0.22	1.00							
Important (4)	0.27	0.14	0.23	1.00						
Keepup (5)	0.19	0.36	0.12	0.09	1.00					
LikeMaths (6)	0.54	0.35	0.50	0.34	0.17	1.00				
Likemore (7)	0.41	0.30	0.45	0.26	0.15	0.65	1.00			
Probfun (8)	0.49	0.34	0.38	0.31	0.21	0.70	0.56	1.00		
Seewhy (9)	0.35	0.27	0.18	0.33	0.22	0.28	0.19	0.25	1.00	
Useless (10)	0.41	0.27	0.22	0.35	0.24	0.39	0.25	0.36	0.43	1.00

The oblique solution suggested a moderate degree of overlapping variance (15 %) between Components I and II ($r = 0.39$), with modest overlap for Components I and II (5 %) and II and III (6 %) ($r = 0.23$; $r = 0.24$, respectively). As the Varimax rotation produced an identical pattern of loadings, and as correlations for two of the three factor pairs were modest, results of the orthogonal rotation are presented here.

Varimax-rotated component loadings and communalities for the 10 HIFAMS items are presented in Table V. Initially, items with loadings of 0.30 or more were retained for interpretation. Based on this criterion, all 10 of the HIFAMS items loaded on one of the three scale components. Two items (Boring and Confusing) showed some degree of cross-loading across Components I and II and between Components I and III (respectively). However, in each case, the loading for one component was substantially larger. As a result, both items were retained and interpreted as part of the higher loaded component.

Table V. Orthogonal component solution for HIFAMS scores

Item	Communality (h^2)	Component I	Component II	Component III
LikeMaths	0.79	0.84	0.27	0.13
Likemore	0.68	0.82	0.08	0.10
Probfun	0.66	0.75	0.25	0.19
Enjoy	0.48	0.69	0.06	0.02
Boring	0.52	0.54	0.37	0.29
Seewhy	0.64	0.08	0.76	0.22
Useless	0.62	0.18	0.73	0.24
Important	0.61	0.25	0.72	–0.15
Keepup	0.70	0.03	0.10	0.83
Confusing	0.62	0.30	0.13	0.72

As shown in Table V, Component I was defined by five of the HIFAMS items (‘I like maths’, ‘I like maths more than my other school subjects’, ‘Doing maths problems is fun’, ‘I enjoy my maths lessons’, and ‘Maths is boring’). As these items seemed to reflect a general enjoyment of the subject, this component was labelled ‘Enjoyment’. Component II was defined by three HIFAMS items: ‘I can’t see why I have to do maths’, ‘Maths is an important subject’, and ‘Maths is a useless subject’. As these items seemed to relate to an appreciation of the importance and value of the subject, this component was labelled ‘Value’.

Component III was defined by only two items: ‘I can’t keep up with the work we do in maths’ and ‘Maths is too confusing’. As these items appeared to reflect students’ evaluations of their own ability to cope with the subject matter, this component was labelled ‘Coping’. From Table II, there was a moderate degree of overlapping variance between scores on these two items ($r = 0.36$). Although correlations between the item ‘I can’t keep up with the work we do in maths’ and other items in the scale were substantially lower than 0.36 (all rs < 0.25), scores on the item ‘Maths is too confusing’ correlated equally well with the items ‘I like maths’ ($r = 0.35$) and ‘Doing maths problems is fun’ ($r = 0.34$). As a result, internal consistencies were also lower for this component ($\alpha = 0.55$) than for the value and enjoyment components ($\alpha = 0.69$ and $\alpha = 0.83$, respectively).

Despite this, the component was retained for several reasons. First, the distinction between students' evaluations of their confidence in, or difficulty of, various school subjects has emerged with the use of several previous attitude scales (eg the SSAS and the FSMAS). Thus, the composition of the component was consistent with previous research in the area. Second, the component appeared to be well defined by the two items, with loadings of 0.82 and 0.72 and low or negligible loadings (all $rs \leq 0.30$) on the other two components. The component was also found to be consistent across all sub-sample analyses described below.

To examine the stability of the three-component solution across the four grade levels, separate PCAs were performed on the scores for students in grades four, five, six and seven. Equivalent Varimax-rotated solutions were reached for the former three subsamples, with each PCA producing three components (eigenvalues greater than one), with equivalent loading patterns for the 10 scale items. Correlations between loadings on the three subscales generated by the full sample and each grade subsample analysis were high ($rs \geq 0.91$, $df = 8$). Correlations were also high between component scores based on the full sample and subsample analyses for grade four ($rs \geq 0.97$, $df = 104$), grade five ($rs \geq 0.98$, $df = 292$) and grade six ($rs \geq 0.98$, $df = 312$).

A three-component solution was also reached for the grade seven subsample. However, the composition of the last two components differed markedly from that of the full sample solution. Component I for the grade seven sample was similar to Component I of the full sample solution (Enjoyment). This component included four items: 'I enjoy doing maths problems', 'I like maths more than my other school subjects', 'I like maths' and 'Doing maths problems is fun'. However, Component II for the grade seven subsample comprised three items ('Maths is too confusing', 'Maths is boring' and 'I can't see why I have to do maths'). Component III also comprised three items ('Maths is an important subject', 'I can't keep up with the work we do in maths', and 'Maths is a useless subject').

As a result, the correlation between the grade seven and full sample loadings for Component I was high ($r = 0.94$), but low correlations were found between the loadings for Components II ($r = 0.07$) and III ($r = -0.05$). The second seventh-grade component correlated moderately with Component III (Coping) of the full sample solution ($r = 0.69$). As the composition of the final two components of the grade seven solution are less intuitively reasonable than those of the full sample solution, it is possible that this outcome reflected the low number of students in the grade seven sample ($n = 60$ as compared with $n = 106$, 294 and 314 for the grade four, five and six samples, respectively). Given this, and the fact that the solution was found to be stable across the other three grade levels, the result for the grade seven sample is unlikely to represent a true moderator effect.

The stability of the solution was also examined across the male and female subsamples. A number of students chose not to indicate their gender on the survey forms. As a result, separate analyses for males and females could be conducted only for 558 subjects (294 females and 264 males). Both subsample analyses indicated that the solution reached for the full sample was robust across males and females. The solution reached for the female subsample was identical to the full sample solution, and component loadings generated by the two analyses were highly correlated ($rs \geq 0.96$, $df = 8$). A high correlation was also found between the component scores generated for female students based on the full sample and female subsample analyses ($rs \geq 0.98$, $df = 292$). Thus, the solution reached for the female subsample was equivalent to the full sample solution.

For the male subsample, a two-component solution was initially reached using Kaiser's latent root criterion. However, when an *a priori* three-component solution was obtained, the pattern and magnitudes of the loadings produced were similar to those produced in the full sample analysis. The eigenvalue for Component III in the male subsample analysis fell only marginally below the Kaiser cutoff (0.94). Correlations between the component loadings produced by the two analyses were high ($rs \geq 0.96$, $df = 8$), as were the correlations between loadings produced by the male and female subsamples ($rs \geq 0.87$, $df = 8$). Correlations between component scores generated by the male and full sample analyses were also high ($rs \geq 0.98$, $df = 262$). Thus, the results produced by the male subsample were also similar to those produced by the full sample analysis.

As a final check on the stability of the obtained three-factor solution, the 774 sets of scores were assigned randomly to one of two groups (stratifying for grade level), and separate PCAs were performed for these two split-half subsamples. These also produced equivalent solutions to the full-sample analysis. Component loadings produced by these analyses were highly correlated (all $rs \geq 0.99$, $df = 8$), indicating a high level of correspondence between the solutions.

Temporal stability

To obtain four-week temporal stability estimates for total and factor scores on the HIFAMS, four classes (one from each of the four grade levels) were randomly selected to complete the HIFAMS again after a four-week period. Unlike in the subsample analyses, where a regression approach was used to estimate component scores (to facilitate the comparison of these scores across subsample solutions), component scores in this case were generated by summing scores for items within each component. The correlations shown in Table VI indicate temporal stabilities for total HIFAMS and component scores across all the four grade levels. As indicated, these were generally within expected ranges for attitudes scales, with only three estimates falling below 0.70. The results do, however, indicate some degree of instability for the Enjoyment component at grade four, and the Value and Coping components at grade six.

Table VI. Four-week temporal stability estimates

Grade	Male N	Female N	Enjoyment	Value	Coping	Total score
4	9	15	0.66	0.70	0.68	0.83
5	15	15	0.81	0.75	0.75	0.86
6	9	15	0.93	0.59	0.59	0.87
7	10	11	0.90	0.77	0.77	0.91
Total	43	56	0.81	0.72	0.70	0.86

Correlations with teacher ratings

Teachers in the four selected classes were also asked to rate, on a five-point scale (ranging from very much so to not at all), their perceptions of how much each of their students liked maths; their perceptions of the value of the subject; and the extent to which they appeared to be coping with their assigned work in maths. Teachers did not have access to their students' HIFAMS responses. These ratings were then correlated with students' scores on each of the three HIFAMS components. Teachers' ratings were also summed to provide overall total scores, which were correlated with students' total HIFAMS scores. These correlations are presented in Table VII. As indicated, all but one of the correlations were positive and above 0.30.

Table VII. Correlations between HIFAMS subscale scores and teacher ratings

Grade	Enjoyment	Value	Coping	Total score
4	0.57	0.48	0.16	0.52
5	0.51	0.55	0.47	0.66
6	0.61	0.33	0.33	0.76
7	0.34	0.64	0.73	0.47

The one correlation that fell below 0.30 ($r = 0.16$) indicated a poor correspondence between the ratings of fourth graders regarding the extent to which they were coping in maths, and their teachers' perceptions. This result may reflect the fact that for the other two subscales, teachers rated what they considered to be the perceptions of their students, rather than their own. For this item, teachers gave a rating of how well students were actually coping with the subject. Thus, this result may suggest that the perceptions of fourth graders regarding their own relative performance may be somewhat less accurate than those of students in higher grade levels.

Conclusion

Results of this preliminary validation generally supported the technical characteristics of the HIFAMS. Analyses of the principal components indicated that the scale items were well represented by a three-component solution, which was found to be stable across grade levels, sex, and a random split-half check. Internal consistencies and temporal stability estimates were moderately high for the Enjoyment and Value components. For the Coping component, both indices were somewhat lower either in the full sample or within grades. This is likely to reflect the relatively low number of items in this subscale. This problem can be addressed with refinements to the scale, by adding further items on the extent to which students feel that they are coping with their assigned work in mathematics.

The overall structure of the scale corresponds well with Nyberg and Clarke's (1979) Evaluation, Usefulness and Difficulty subscales. Components I and II of the present scale (Enjoyment and Value) were also similar in content to the first two components of the Enjoyment and Value subscales of the Aiken (1974) Mathematics Attitude Scales, while Components II and III (Value and Coping) resembled the Usefulness and Confidence subscales of the Fennema-Sherman Mathematics Attitude Scales (Fennema & Sherman 1976). Thus, the results of this study provide further support for the existence of these three dimensions in students' attitudes to school subjects.

Due to its relative brevity and the reduced complexity of item wording, however, the present scale is suitable for use with students in the primary grades. With each administration, students read and completed the survey autonomously, but were instructed to seek clarification for any ambiguous scale items. In these sessions, the types of questions asked were generally focused on matters of confidentiality and identification numbers, rather than about confusion over the scale content. Support teachers working with students who had reading difficulties also reported that the items were easily understood by these students. In general, the scale took less than 10 minutes to complete. Thus, the HIFAMS holds promise for providing researchers and educators with a brief measure of attitudes to mathematics that is valid and suitable for use with students in the primary grades.

References

American Association of University Women (1992) *Shortchanging girls, shortchanging America: a call to action.* ERIC Document Reproduction Service No. ED 340658.

Aiken LR (1974) Two scales of attitude towards mathematics. *Arithmetic Teacher*, vol 19, pp 229–234.

Aiken LR & Dreger RM (1961) The effect of attitudes on performance in learning mathematics. *Journal of Educational Psychology*, vol 52, pp 19–24.

Australian Bureau of Statistics (1997) *1996 Census of Population and Housing.* Canberra, Australian Capital Territory: The Australian Bureau of Statistics.

Dungan JF & Thurlow GR (1989) Students' attitudes towards mathematics: a review of the literature. *The Australian Mathematics Teacher,* vol 45, no 3, pp 8–11.

Fennema E (1996) Mathematics, gender, and research. In G Hanna (ed) *Toward gender equity in mathematics education.* Boston: Kluwer.

Fennema E & Sherman JA (1976) Fennema-Sherman Mathematics Attitudes Scales: instruments designed to measure attitudes towards the learning of mathematics by males and females. *JSAS Catalog of Selected Documents in Psychology,* vol 6, no 1, p 3b.

Frost LA, Hyde JS & Fennema E (1994) Gender, mathematics performance, and mathematics-related attitudes and affect: a meta-analytic synthesis. *International Journal of Educational Research,* vol 21, pp 373–386.

Greenwald AG & Banaji MR (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review,* vol 102, no 1, pp 4–27.

Hyde JS, Fennema E, Ryan M, Frost LA & Hopp C (1990) Gender comparisons of mathematics attitudes and affect: a meta-analysis. *Psychology of Women Quarterly,* vol 14, pp 299–324.

Ma X & Kishnor H (1997) Assessing the relationship between attitudes towards mathematics and achievement in mathematics: a meta-analysis. *Journal for Research in Mathematics Education,* vol 28, no 1, pp 26–47.

Melancon JG, 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,* vol 54, no 1, pp 187–192.

Mulhern F & Rae G (1998) Development of a shortened form of the Fennema-Sherman Mathematics Attitudes Scales. *Educational and Psychological Measurement,* vol 58, no 2, pp 295–306.

Neale DC (1969) The role of attitudes in learning mathematics. *Arithmetic Teacher,* vol 16, pp 631–640.

Nyberg VR & Clarke SC (1979) *Technical report on the School Subjects Attitude Scales.* ERIC Document Reproduction Service No 179 575.