Development and validation of two learning outcome measures: 'Disposition to enter HE' and 'disposition to study mathematically demanding subjects in HE'

Julian Williams, Maria Pampaka, Laura Black, Pauline Davis, Paul Hernandez-Martinez, Geoff Wake and Julian Williams
School of Education,
University of Manchester

Abstract

The aim of this paper is to describe and validate the development of two measures constructed to measure AS students disposition (i) to enter HE and (ii) to further study mathematically-demanding subjects, which we regard as potentially significant variables in monitoring or even explaining students progress in to different studies in HE. The items for the scale were constructed on the basis of interview data, and drew on a model of disposition as socially- as well as self- attributed. Drawing on Rasch analyses of pilot and 'main' data sets, we find that the two scales each produce healthy one-dimensional fits on what we take to be a 'strength of commitment to enter HE' and 'disposition to study mathematically-demanding subjects further' respectively. However, as a measurement scale for this sample in this context the former scale suffers from a ceiling effect: our sample are overwhelmingly committed to entering HE (at the early stages of AS level mathematics course anyway). To 'correct' for this, we added some harder items to the analysis at a later data point, and found (i) an item that improved separability of the instrument for the higher scorers, and (ii) a massively misfitting, hard item that is worthy of future research.

1. Introduction

This paper is based on quantitative analyses of a data-set from the ESRC TLRP research project on widening participation in HE, 'Keeping open the door to mathematically-demanding F&HE programmes'. We particularly draw on sets of responses to a questionnaire administered to students undertaking AS Maths and AS Use of Maths courses, and focus on the development and validation of two new instruments for measures of students disposition towards HE and further studying mathematically demanding subjects.

The economic significance of mathematics and the shortage of mathematically well-qualified students and graduates (i.e. the 'Mathematics Problem') is strongly emphasised by recent reports (i.e. Smith, 2004). Hence, we need to understand how mathematics can become more accessible to students, especially those students for whom AS/A2 mathematics is a barrier to progressing into mathematically demanding courses that confer social, cultural and economic capital. The general aim of the research project is to understand how to widen participation in mathematically demanding subjects generally, but particularly for our 'target' students, i.e. those students who are at the margins of continuing with maths.

Towards this end we encountered the need for measures of students' 'perception of intention to study in HE', and additionally measures of their intention to persist in study of mathematically-related topics in Further and Higher Education. To our knowledge there is as yet no existing measure of intention to persist in the study of mathematically related topics in F&HE. Nor is there a general measure of intention to enter HE that has been validated in F&HE that meets our requirements. However, there is an eclectic but relevant literature that informs the development and

validation of such educationally socio-culturally sensitive measures (i.e. Eley and Meyer, 2004; Hoyles, et al, 2001).

The first instrument, namely 'disposition to enter HE' consists of four statements eliciting students own expectation about themselves going to university and the expectations of others about this possibility (family, friends, teachers). Our view, supported by interviews, was that students might be said to exhibit a stronger commitment to HE if they said that significant others had such expectations for them (though in fact we noticed that they were sometimes less sure of their teachers' views than they were of their family and friends!)

The second instrument consists of 6 items aiming to capture information on students' dispositions to studying mathematically demanding subjects in future HE. The items included in both instruments are presented to students in a multiple choice format and have various numbers of response categories. This had direct implications for the selection of the appropriate (partial credit) measurement model to be selected when calibrating these instruments. Validation was performed by employing the Rasch Partial Credit Model (Bond & Fox, 2001) on a pilot sample of the project (N=314) and suggested robust measures. Some problems appeared regarding the HE disposition instrument, because of sample characteristics, i.e. high tendency of the particular group to report a disposition of "going to HE", and we tried to overcome these with additional items in the main study sample. We will report on both these results in this paper.

We plan to use these validated soft measures as explanatory variables for exploring the effectiveness of different FE maths programmes, as well as a predictor of students' future decisions / choices at UCAS, in the next paper (Pampaka et al., this symposium). We finally discuss the possibility that these instruments will have utility in the wider widening participation research community.

2. The Rasch Partial Credit Model

George Rasch, realized that, to be of any use at all in a measurement model, a measure must retain its quantitative status, within reason, regardless of the context in which it occurs. Like a yardstick, each test or any other construct's item must maintain its level of difficulty, regardless of who is responding to it. It also follows that the person measured must retain the same level of competence or ability regardless of which particular test items are encountered, so long as whatever items are used belong to the calibrated set of items which define the variable under study. Rasch also recognized that the outcome of an interaction between an object-to-be-measured, such as a person, and a measuring-agent, such as a test item, cannot, in practice, be fully predetermined but must involve an additional, unavoidably unpredictable, component (Wright & Linacre, 1989).
The Rasch model (named after its developer) provides the means for constructing interval measures from raw data. When data can be selected and organized to fit a Rasch Model, the cancellation axiom of additive conjoint measurement is satisfied, a perfect Guttman order of response probabilities and hence of item and person parameters is established, and items are calibrated and persons measured on a common interval scale. The model proposes a mathematical relationship between a person's ability, the difficulty of the task, and the probability of the person succeeding on that task (Wright & Mok, 2000; Acton, 2003; Wright, 1999).
The family of Rasch models is the only one that solves measurement problems, because these models produce linear measures, overcome missing data, give estimates of precision, have devices for detecting misfit and the parameters of the object being measured and of the measurement instrument are separable (Wright and Mok, 2000). As Masters also (undated) notes a unique feature

of the Rasch model is that "when data fit the Rasch Model, it is possible to compare abilities without knowing, or even having to estimate, the difficulties of the task" (p. 23).
Applications of Rasch Measurement, among others, include Item banking, test design, tailored testing, self-tailoring, response validation (through the analysis of fit) and item bias (Wright, 1999) which can be realized with different statistical analyses. When the name 'Rasch Model' appears in connection with statistical analyses, it often means the so-called 'dichotomous Rasch Model' or the 'one-parameter logistic model' which is the simplest format of the models and records only two levels of performance on an item, e.g. 'Fail/Pass'.
The Partial Credit Model (PCM), which is used for the purposes of this paper, is considered to be an extension of the simple dichotomous (Rasch) model, for outcomes recorded in more than two ordered response categories, i.e. by awarding 'partial credit' for responses that are neither correct nor totally incorrect. The model can be applied to any set of test or questionnaire data collected for the purposes of measuring abilities, achievements, or attitudes provided that responses to each test or questionnaire item are scored in two or more ordered categories (Masters, 1982,1999). Bode (2001) lists three specific situations in which the PCM can be used, and those are when instruments contain items:
with varying degrees of correctness for responses that can be ordered from least correct to most correct,
that can be broken into component tasks, the first of which must be completed before the next is attempted, and each of which can be scored as correct or incorrect,
where increments in the quality of performance are rated.

Athanasou and Lamprianou (2002) clarify what is additionally needed (compared to the simple model) in order to draw the graphs for PCM (with the score probability lines for a question/item): whereas one parameter for the people (the ability) and one parameter for the items (the difficulty of getting 1 instead of 0) is enough for the simple model, in the partial credit model additional parameters should be introduced as the δ (difficulty) parameter cannot describe the question fully. It is essential to know the difficulty of achieving each of the score categories.

The PCM specifies that each item has its own rating scale structure, that is, the transition from one category to the next can have a different meaning from one item to the next (Wright, 1999; Bode, 2001). In the following sections we will show how this model was employed to validate the two measures of disposition.

3. Results

The subjects (sample description and data collection)

Results presented in this paper are based on analysis of data from (i) a pilot study, and (ii) the first two data points of the longitudinal, main stages of the project (DP1 and DP2) as shown below in BOLD under 'dispositions'. Note that these two instruments are called HEdisp and MathDis: the other disposition instrument is a Maths self-efficacy instrument that we have reported elsewhere.

Table 1: Design of data collection

		Year 1 (2006-7)		Year 2 (2007-8)
Background Variables		DP 1 Start of AS	DP2 End of AS	DP3 Start of A2
Family-in-HE	Hard LOs	GCSE grades	AS grades	

LPN by postcode, EMA, Gender Language(Eng/ non-Eng/bi), Ethnicity, College	**Dispositions:** **HE Dis** **MathDis** **MSE**	**HEDis-1** **MathDis-1** **MSE-1**	**HEDis -2** **MathDis-2** **MSE-2**	**HEDis -3** **MathsDis-3** **MSE-3**
	Intentions/ Choices/ Decisions	Uni-int 1 STEMint1	Uni-int 2 STEMint2	Uni-int 3 STEMint3

Pilot Data came from 313 AS student from 23 different further education institutions in UK (their distribution regarding gender and course is shown in Table 2). 27 GCSE students (i.e. students who have not yet started an AS course: 15 male, 12 female) were additionally involved in the pilot study). Table 2 also shows the distribution of the students at the next two main stages of the study (Data point 1 and Data point 2)

Table 2: Distribution (frequencies) of students according to gender and course

	Maths Course		Total
Gender	**AS Trad**	**AS UoM**	
PILOT			
Male	144	55	199
Female	70	44	114
Total Pilot	**214**	**99**	**313**
DP1			
Male	769	341	1110
Female	511	153	664
Total DP1	**1280**	**494**	**1774**
DP2			
Male	413	235	648
Female	288	108	396
Total DP2	**701**	**343**	**1044**

'Disposition to study mathematically demanding HE courses'

We begin with the results for the 'disposition to study mathematically demanding subjects in HE' scale, as this proved least problematic form a measurement point of view. Table 3 shows the items and the pilot frequencies of responses: these items were informed by our pilot interviews with students, and were all typical of the kinds of things students said to us in informal conversations and interviews about their intentions and dispositions.

Table3: Items in the 'Disposition to study mathematically demanding courses in HE' instrument, with pilot frequencies

PC coding	Item and Responses	ITEM NAME Frequency (pilot)
	Are you planning to study any more mathematics courses or units after this AS course? [B1]	PLAN
2	Yes	197
0	No	82
1	Don't know	69
	My preferred options for a course at university will include: [B8]	AMOUNT
4	A lot of mathematics	38
3	Quite a lot of mathematics	72
2	A moderate amount of mathematics	100
1	As little mathematics as possible	40
0	No mathematics	21
9	Don't know	53
		32 Missing
	The amount of mathematics in my preferred options for the course at university was: [B9]	IMPORTANCE
2	Very important	53
1	Quite important	131
0	Not at all important	76
9	Don't know	62
		Missing 32
	If I find out that my future course involves studying more mathematics than I thought, this would make me feel: [B10]	FEELINGS
4	Very happy	30
3	Fairly happy	92
2	Indifferent	77
1	Unhappy	13
0	Very unhappy	70
9	Don't know	44
		30 missing
	If in the future I am studying a course involving mathematics, then I would prefer it to be: [B11]	MATH TYPE
0	Familiar mathematics that I have already done	98
2	New mathematics that I have not learnt before	22
1	A mix of familiar and new mathematics	184
	Don't know	62

Analysis and Results

Analysis was conducted on the full data sets for Data point 1 and 2 using a Partial Credit Model, in the software 'QUEST'. The fit statistics are better than acceptable for a case such as this, as shown below in Table 4a,b for the pilot data and the main data points 1 and 2:

Table 4a: Measures (Thresholds) and fit statistics for the items of the MHE-disposition Scale at Pilot

Item name (number)	Thresholds (SE)				Infit MNSQ	Outfit MNSQ
	1	2	3	4		
Plan (1)	-1.25 (.25)	-0.27 (.26)			1.06	1.13
Amount (7)	-2.44 (.38)	-1.21 (.29)	0.41 (.24)	1.95 (.31)	0.78	0.78
Import (8)	-0.94 (.31)	1.61 (.32)			0.88	0.88
Feelings (9)	-3.00 (.41)	-0.77 (.26)	0.26 (.26)	2.32 (.34)	1.17	1.17
MathType(10)	-0.88 (.28)	2.90 (.40)			1.08	1.08
Mean	-0.01				0.99	1.01
SD	0.69				0.16	0.17

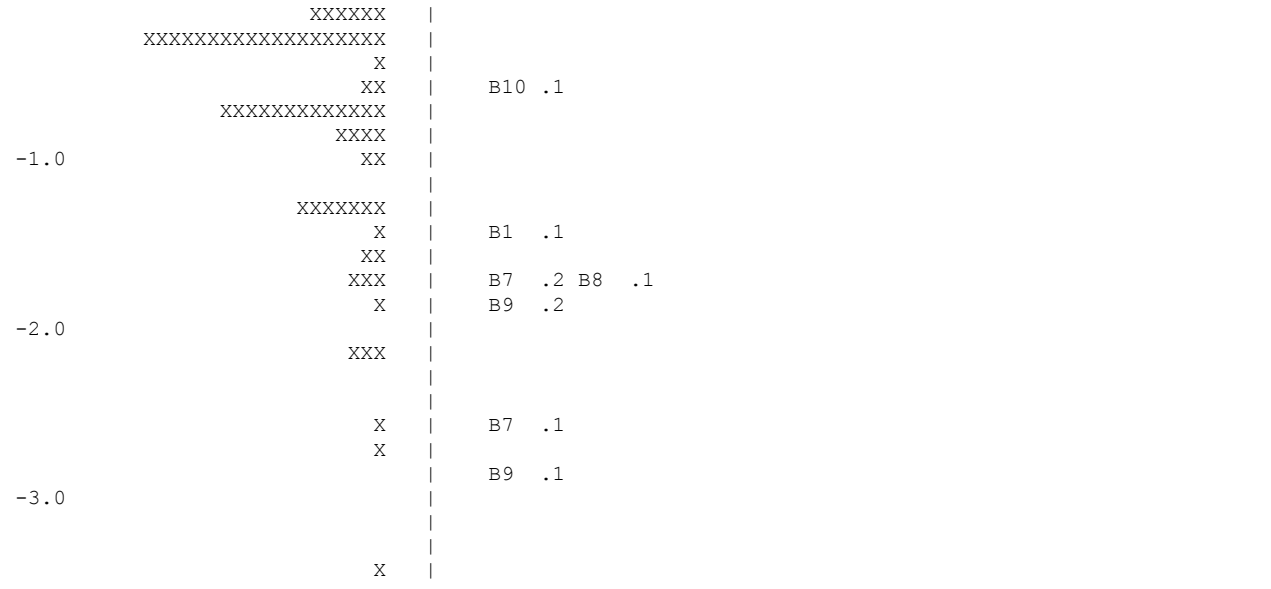
Table 4b: Fit statistics at DP1 and DP2

Item	Infit MSQR		Outfit MSQR	
	DP1	DP2	DP1	DP2
Plan	0.95	0.93	0.98	0.97
Amount	0.80	0.81	0.79	0.82
Importance	0.91	0.99	0.90	1.02
Feelings	0.95	0.91	0.96	0.94
Math Type	1.31	1.27	1.37	1.34
Mean	0.98	0.98	1.00	1.02
SD	0.19	0.17	0.22	0.19

The item map (Figure 1) for data point 1 shows how the students scores are separated by the items in a very efficient way, any improvement in the measure would be at the cost of adding extra items or refinements of codings, in order to reduce errors of estimation, probably n the middle of the spread.

Figure 1 : Item map for HE maths instrument at data point one

```
-------------------------------------------------------------------------------------------------
Each X = 12 students     Item Estimates (Thresholds)  (number of item given in table 4
-------------------------------------------------------------------------------------------------
|
                            XX    |
                                  |
                                  |
                                  |
  3.0                             |    B10 .2 (Math type)
                                  |
                           XXXX   |
                                  |    B7  .4 (Amount) B9  .4 (feelings)
                                  |
                                  |
                                  |
  2.0                  XXXXXXXX  |
                               X  |
                                  |
                  XXXXXXXXXXXXX   |
                               X  |
                               X  |    B8  .2 (importance) B9  .3
                                  |
  1.0           XXXXXXXXXXXXXXXX  |    B7  .3
                               X  |
                              XX  |
                             XXX  |
           XXXXXXXXXXXXXXXXXXX   |
                                  |    B1  .2
  0.0                             |
```



The HE disposition measure

However, the story with the other measure was more problematic. The HE disposition scale consisted of the following 4 items/response codes: see table 5

Table 5: Information for the items of the scale 'Disposition to enter HE'

Item description	Item name	Question and Responses	PC Scoring	Freq PILOT	Freq DP1
My expectation	"self"	Are you planning to go to University? Yes Depends No	 2 1 0	 261 78 14	 1207 440 64
Family Expectation	"family"	(My family expects that) That I will go to university Different/conflicting expectations **Indifferent ????** That I will not go	 2 1 1 0	 272 16 37 9	 1212 470 31
Friends Expectation	"friends"	(My friends expect that) That I will go to university Different/conflicting expectations NO expectation Don't know That I will not go	 2 1 1 1 0	 271 28 23 10	 1013 655 44

Teachers' expectation	"teachers"	(My Teachers expect that) That I will go to university Different/conflicting expectations NO expectation Don't know That I will not go	 2 1 1 1 0	 264 24 13 11	 852 844 10

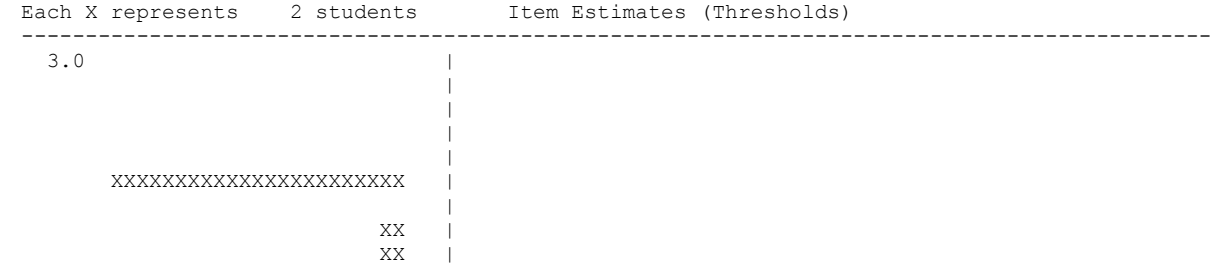
Analysis here involved the use of the Partial Credit Model, for outcomes recorded in more than two ordered response categories, i.e. by awarding 'partial credit' for responses that are neither correct nor totally incorrect (Masters, 1999). In this case, partial credit was awarded in items with 'intermediate' levels of expectancy to go to university, e.g. the responses coded as '1' in Table 2. The results of the pilot analysis are shown below. As shown in Table 3 item fit analysis seems adequate: all items infit statistics are within the acceptable limits expected in this kind of research. However, the fit of the' teachers' item in DP1 is interesting: the relatively high Outfir suggests that there are at least some students whose responses tot his item do not fit as well as the others. This might be explained by the fact that the se students get less feedback form their teachers on this issue than they do from friends and family, or that the feedback they get might be dissonant with their own view of themselves. The current project does not have ambitions to further explore this question, but the data are somewhat suggestive.

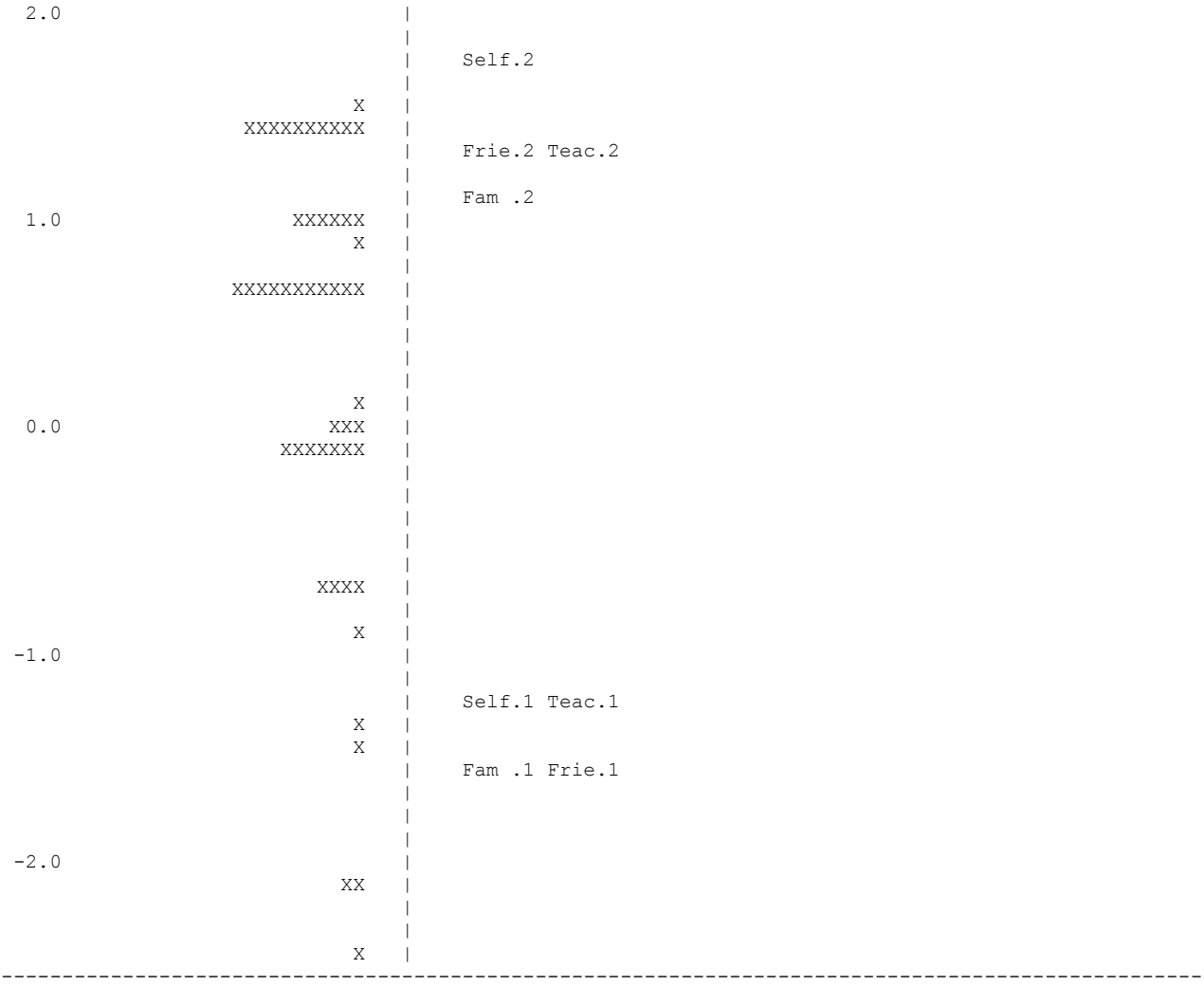
Table 6: Measures (Thresholds) and fit statistics for the items of the HE-disposition Scale at Pilot and Data Point 1

	Thresholds (SE)				**Infit MSQR**		**Outfit MSQR**	
	Pilot		**DP1**		Pilot	DP1	Pilot	DP1
Item	1	2	1	2				
Self	-1.25 (.50)	1.75 (.37)	-1.72 (.25)	2.08 (.15)	1.04	0.90	1.03	0.90
Family	-1.56 (.59)	1.16 (.42)	-2.69 (.31)	2.07 (.13)	0.92	0.93	0.92	0.86
Friends	-1.53 (.56)	1.34 (.39)	-2.25 (.31)	2.96 (.15)	1.00	0.85	1.00	0.81
Teachers	-1.25 (.56)	1.32 (.41)	-4.13 (.47)	3.67 (.14)	1.01	**1.20**	0.99	**1.87**
Mean	0.00		0.00		0.99	0.97	0.98	1.11
SD	0.19		0.32		0.05	0.16	0.05	0.51

Figure 2 presents the actual maps of items difficulties and person's ability in the common constructed scale for the pilot data and DP1 analysis:

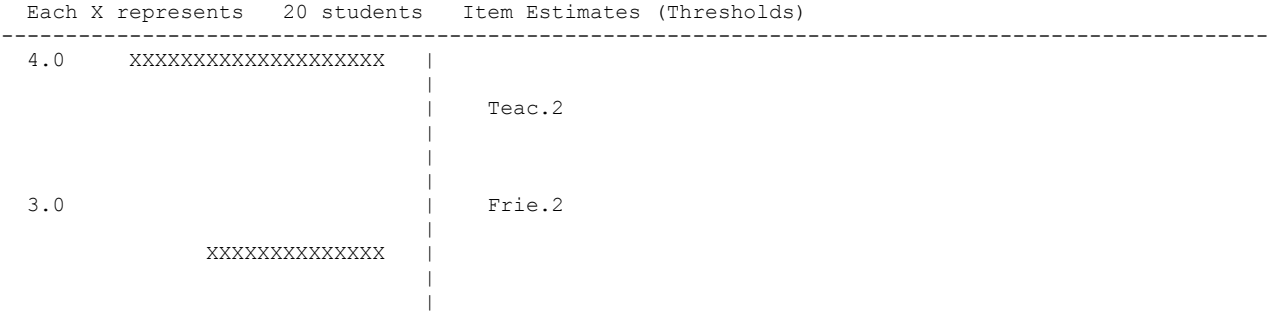
Figure 2a: The "HE Disposition" scale (pilot data) item map and fit

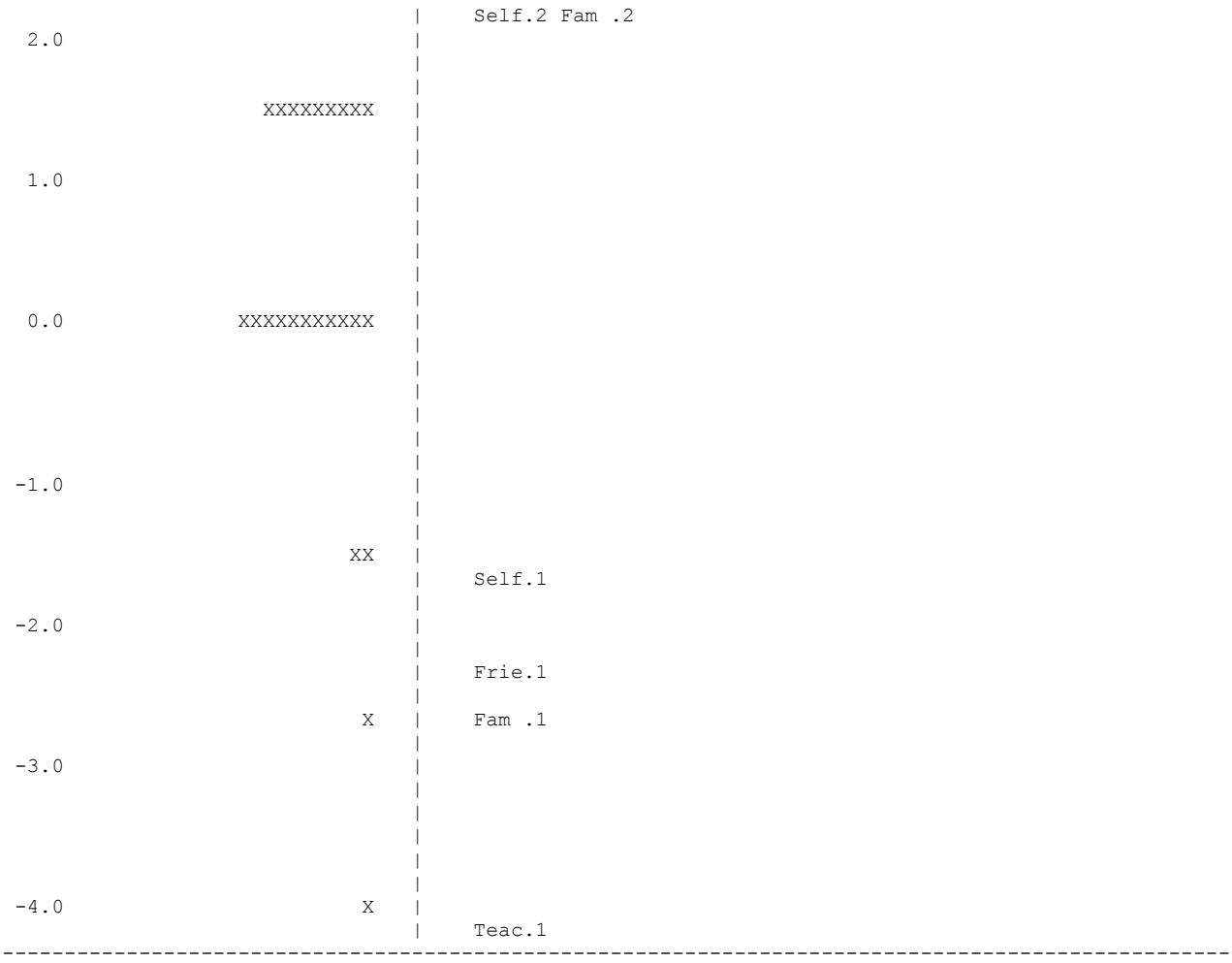




An observation of the pilot results concerned the fact that the scale could not discriminate a lot for the students of this sample since most of them appear to be gathered at the top end of the scale, meaning that they have high disposition to enter HE. (It should also be noted that many students appeared with perfect scores in this scale and their actual logit measure could therefore not be calculated without making sample-assumptions). This was believed to be related to the timing of the pilot at the end of AS year, and we expected to have more spread of responses at the beginning fo the College course, which was the relevant moment for the main stage of the study. Therefore, the decision was taken for this instrument to remain as it is and no changes to be made, apart from some minor modification in the wording of the items to make them consistent.
The story of the measure during the next stage is shown in following figure, 2b.

Figure 2b: The "HE Disposition" scale (DP1 analysis) item map





One notices in both these item maps some lack of separation and discrimination in the middle of the scale, with many item thresholds grouped. This is suggestive of a ‘levelling’ in the sample, and one might be able t establish some criteria for developing a hierarchy.

More important and problematic however, it seems that, contrary to our expectation, the measure remained unable to discriminate the high end of the spectrum, which includes about 400 students in our sample, who still get the highest score. It is too “easy” for our target groups. Hence the decision was taken to try to extend the measure to include a little bit more difficulty by adding two more difficult items at DP2. (See table 7).

Considering a new HE measure at DP2

Two new Likert type items were included in the instrument to improve the separability of the scale These are in table 7, and the fit statistics (a surprise!) are in Table 8: the model seems to indicate a bad fit to "Repeat" on both infit and outfit, suggesting that this misfit is not caused by a few off-target responses but is probably an issue for students right across the measured spectrum.

The infit for the new item "Take it" is healthy: while the item has a slightly higher infit than the others, it is also a hard item, and it is suggestive that this infit is related to the fact that a hard item on a scale is vulnerable to a few off-target residuals. In this case the relatively high outfit for "Take it" (1.45) is consistent with this interpretation. We therefore conclude that the measure benefits form "take it" while "Repeat" potentially threatens the construct as a whole.

Table 7: two additional items for the HE disposition scale, "Repeat" and "take it"
:

	Disagree strongly	**Disagree**	**Agree**	**Agree strongly**
New1: **Repeat** " I am prepared to repeat a year at college in order to get into university, if necessary."	1	2	3	4
New2: **Take it "** If I was offered the career I wanted without having to go to university, I would consider taking it. " NB: this item is Reversed coded	4	3	2	1

Table 8a: the fit statistics for the HE instrument at data point 2

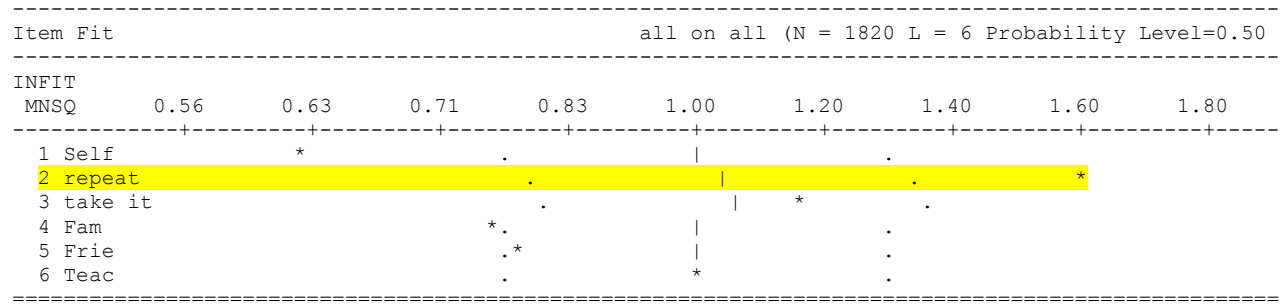
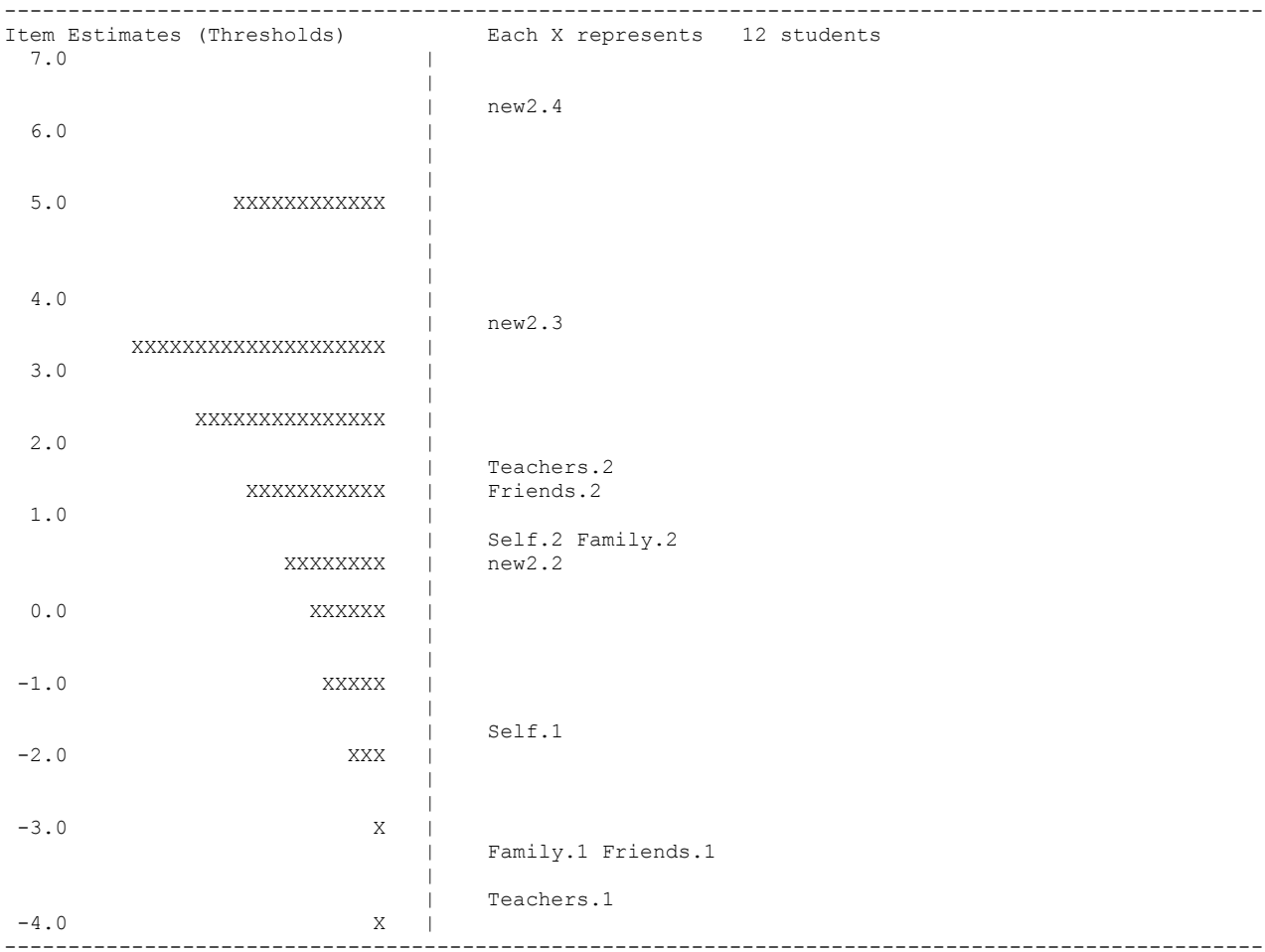


Table 8b: Fit statistics for analyses on different data sets for the HE disposition with and without the 2 new items

Item	**Infit MSQR**				**Outfit MSQR**			
	DP1	DP2-A	DP2-B	DP2-C	DP1	DP2-A	DP2-B	DP2-C
Self	0.90	0.86	0.62	0.75	0.90	0.88	0.58	0.84
Repeat			**1.56**				**1.62**	
Take it			**1.10**	**1.19**			**1.16**	**1.45**
Family	0.93	0.88	0.76	0.83	0.86	0.83	0.66	1.07
Friends	0.85	0.89	0.78	0.86	0.81	0.81	0.68	1.01
Teachers	1.20	1.22	0.99	1.13	1.87	1.48	0.97	1.47
Mean	0.97	0.96	0.97	0.95	1.11	1.00	0.94	1.17
SD	0.16	0.17	0.34	0.19	0.51	0.32	0.40	0.28

The decision to exclude the item from the measurement scale does have the effect of reducing the difficulty of the scale as a whole however, and leads in the end to us having to deal with numbers of cases of students with 'full score' still. Nevertheless the final scale including the new item has better separability at the top end of the scale than the original scale, as is indicated in Figure 3: notice that the responses '3' and '4' on this item are particularly critical!

Figure 3: Item map for HE disp - Analysis of Data Point 2- with the new item "Take it"



The item "Repeat" raises another issue however, why does this item misfit so badly? We assume that it taps into some other dimension, that there is a significant sample sub-group perhaps that is responding to the item in ways that are distinct because of their particular situations. We speculate that:

- there may be some students for whom staying another year at College, for specific reasons of their teachers, peer group, or their grant status, might find staying on unusually acceptable or unacceptable;
- there may be some relation between personal self-esteem and staying on that interferes particularly with certain social groups.

This is an empirical question and worthy of further research involving examination of the item parameters for this item in various subgroups: we are testing for different background variable effects and may be able to report this soon. First analyses suggest that Gender, Programme (use of maths versus Traditional Mathematics) and class proxy variables are implicated in differential item functioning.

Conclusions

We have constructed 'robust' measures of students' dispositions (i) of their disposition to study further mathematically-demanding subjects in HE, and (ii) of their commitment to study in HE, based on short multiple choice items given to students studying AS level mathematics or Uses of mathematics.

The first of these scales provides good separation of the sample, while the second proves problematic at the top end: this led us to extend the scale with one harder item that fits tolerably, and reduces the tendency to 'total scores' on the scale.

An interesting item that did not fit was also analysed and rejected as a possible addition: this leads us to some further investigation of the construct, especially as concerns class, gender and other aspects of background.

Discussion

We speculate now as to the possible use of this scale for other researchers' purposes. The scalability of an instrument like this is of course an empirical question: will the scale work for other sample sin other situations? It may be that our group of students, predominantly following AS level courses (though with some significant numbers of BTEC) have relative high commitments to HE entry, and that the sale may actually be more suited to the population as a whole than to our own sample. However, our own sample – chosen to reflect a group of students whose hold on education is relatively tenuous - is relatively skewed to lower social classes in urban areas, and so this might also be interpreted appropriately.

References

Bond T.G. and Fox, C.M. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. Lawrence Erlbaum assoc: Mahwah, NJ.

Eley, M. G., & Meyer, J.H.F. (2004). Modelling the Influences on Learning Outcomes of Study processes in University Mathematics, *Higher Education*, 47: 437-454.

Hoyles, C., Newman, K. and Noss, R. (2001). Changing patterns of transition from school to university mathematics, *International Journal of Mathematical Education in Science and Technology*, 32(6): 829-845.

Smith, A. (2004). *Making Mathematics Count*. HM Stationery Office, London.