

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



**CC creative commons**  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

**BY:** **Attribution.** You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>



Teaching Mathematics  
and Its Applications:  
International Journal of the IMA

### Using Rasch Analysis to identify uncharacteristic responses to undergraduate assessments

Journal:	<i>Teaching Mathematics and Its Applications: International Journal of the IMA</i>
Manuscript ID:	TEAMAT-2009-056
Manuscript Type:	Original Article (Section A)
Date Submitted by the Author:	22-Oct-2009
Complete List of Authors:	Edwards, Antony; Mathematics Education Centre, Loughborough University Alcock, Lara; Mathematics Education Centre, Loughborough University
Keywords:	Undergraduate Mathematics, Analysis, Rasch, Examples



# Using Rasch Analysis to identify uncharacteristic responses to undergraduate assessments

---

ANTONY EDWARDS<sup>†</sup>

Mathematics Education Centre, Loughborough University, UK

LARA ALCOCK

Mathematics Education Centre, Loughborough University, UK

**Rasch Analysis is a statistical technique that is commonly used to analyse both test data and Likert survey data, to construct and evaluate question item banks, and to evaluate development in longitudinal studies. In this paper we introduce the dichotomous Rasch Model, briefly discussing its assumptions. Then, using data collected in an example-generation task with undergraduate mathematics students, we use Rasch Analysis to (1) rank the task questions and consider this ranking in relation to students' conceptions and (2) identify those students whose responses misfit the model (e.g. students who score highly but get easier questions wrong). We conclude that because these responses misfit the Rasch Model they are worthy of further (qualitative) examination, for instance to aid the discovery and classification of students' misconceptions relative to formal theory.**

## 1. Introduction

Tests, surveys and questionnaires are a powerful method of data collection, much used by both teachers and researchers. In teaching we often want to identify which students understand material, which students are struggling, and in which ways they are struggling. In education systems more broadly we want to ensure that standards are maintained, and that tests are as objective and fair as possible. As researchers in mathematics education we may use these instruments as exploratory tools rather than assessment tools, but we still want to reduce bias and be as objective as possible. We often conduct studies where a small subset of the total population of students is questioned, but we wish to make generalizations to the wider group.

Such tests may be designed, implemented and analysed for a variety of different purposes, but when statistics are employed we classically represent a student's achievement in a test by their total score. However this is not always desirable: given two students scoring 9/10 in a test, both can be represented by the total score of 90%. Let us assume however, that one student answered a question incorrectly that many other students also answered incorrectly, whereas the other student answered a question incorrectly that was answered correctly by most of the other students. Is 90% still a suitable statistic to describe each student's performance? Depending on the purpose of the test, we may be more interested in either type of student, and yet classically we have not distinguished between them.

---

<sup>†</sup> Email: a.w.edwards@lboro.ac.uk

1  
2  
3 The Rasch Model and its application to data, Rasch Analysis, attempt to take these  
4 considerations into account. Like other Item Response Theory (IRT) models, the aim of  
5 Rasch Analysis is to rank questions by difficulty and to rank participants' responses to these  
6 questions. The Rasch Model one of the more commonly used models in IRT (Baker & Kim,  
7 2004), and this is for a variety of reasons, not least that it permits persons to be compared  
8 who may have not answered the same set of questions.  
9  
10

11 The use of IRT models has recently increased in popularity due to the introduction of  
12 affordable technology capable of quickly running the routines (Kline, 2005 p.107), and Rasch  
13 Analysis has been used as the basis of analysis for a diverse range of studies. These  
14 include the analysis of Likert-scale questionnaires to determine course satisfaction (Waugh,  
15 1998), and attainment (Haines & Crouch, 2001). It has been used both to analyse (Ryan &  
16 McCrae, 2006), to construct (Chen *et al.*, 2005), and to use (Dubby & Duckworth, 1979)  
17 banks of test items.  
18  
19  
20

21 Within mathematics education, the Rasch Model is 'being used increasingly as a research  
22 tool by "mainstream" researchers rather than merely by the sophisticated psychometricians  
23 involved in large-scale achievement testing' (Callingham & Bond, 2006). This quote was  
24 taken from the editorial of a special issue on Rasch Analysis of the *Mathematics Education  
25 Research Journal*. Articles in the journal included Watson *et al.*'s (2006) longitudinal study of  
26 students' understanding of chance and probability, and Bradley *et al.*'s (2006) use of Rasch  
27 Analysis to explore Likert data of student's conceptualizations of quality mathematics  
28 instruction. Other articles found their data was unsuitable for Rasch Analysis (Stacey &  
29 Steinle, 2006), or combined Rasch Analysis with other statistical techniques, such as Factor  
30 Analysis (Grimbeek and Nisbet, 2006).  
31  
32  
33  
34

35 In this paper, we introduce the model (in Section 2), and our research task (Section 3). Then  
36 in Section 4 we then use the model to find which of our task questions are more difficult (and  
37 by how much), and discuss the implications of this. Finally in what we believe to be a novel  
38 use of the procedure in mathematics education, we take answers that the analysis identifies  
39 as uncharacteristic and qualitatively explore them. Other authors (Emons *et al.*, 2005, Reise,  
40 2000) have also explored 'person-fit' statistics, but via further statistical techniques and not a  
41 qualitative examination.  
42  
43  
44

## 45 46 **2. The Rasch Model**

### 47 48 **2.1. Assumptions**

49 The central idea of IRT models is that when a person is faced with an item in a test they  
50 have a certain probability of answering that item correctly. The dichotomous Rasch Model,  
51 which we use below, makes the further assumption that the probability of such a person  
52 answering a particular item correctly is a function of the person's ability and the item's  
53 difficulty alone (as opposed to other models which may take into account other variables  
54 such as the likelihood of a person guessing).  
55  
56  
57

58 Immediately it can be seen that one must be careful when claiming that an individual has a  
59 trait called "ability" that is somehow modelled by the Rasch Analysis (and similarly that an  
60 item has a "difficulty" that can be represented by one number alone). The use of the word  
ability as a variable describing how a specific person has performed in a test is a formal

concept in IRT literature, distinct from any semantic-psychological analysis of the terms everyday usage (Rasch, 1980, p.110). It is an assumption of Rasch Analysis that the test is designed to examine only one facet of a person's understanding, and it is this facet that we label "ability," and that we attempt to estimate during Rasch Analysis.

A further assumption of Rasch Analysis is local (stochastic) independence of items, that is, that the probability of answering one item correctly should not be dependent on the answers to other questions. This assumption means that the calculations that must be made are possible using a desktop computer. Rasch Analysis is therefore not suitable for tests where the questions are structured into parts, the latter parts relying on the earlier ones (although there are ways around this). For a full, detailed discussion of the assumptions of Rasch Analysis see Fischer and Molenaar (1995).

## 2.2. The Rasch Model

With the assumptions as above, Rasch Analysis assumes that for each person taking the test there is a parameter measuring their ability, and for each item on the test there is a parameter measuring the item's difficulty. The notation used for these parameters differs from text to text (Fischer and Molenaar (1995) use  $v_j, \beta_i$  whereas Baker and Kim (2004) use  $\eta_j, \delta_i$ ). In this paper we will use the simpler  $D$  for difficulty of an item and  $A$  for ability of a person i.e.:

$$D_1, D_2, D_3, \dots, D_j$$

where  $D_i$  is the difficulty parameter for item  $i = 1, 2, 3, \dots, j$  and

$$A_1, A_2, A_3, \dots, A_q$$

where  $A_p$  is the ability parameter for person  $p = 1, 2, 3, \dots, q$ .

Given a particular item and person  $\{i, p\}$ , the Rasch Model gives the probability that person  $p$  answers item  $i$  correctly by application of equation (1.1). The shape of this graph is drawn in Fig. 1. Notice that the horizontal axis is a measure of the difference between the  $A_p$  and  $D_i$ , rather than just one of these quantities.

$$P(x_{i,p} = 1 | D_i, A_p) = \frac{\exp(A_p - D_i)}{1 + \exp(A_p - D_i)} \quad (1.1)$$

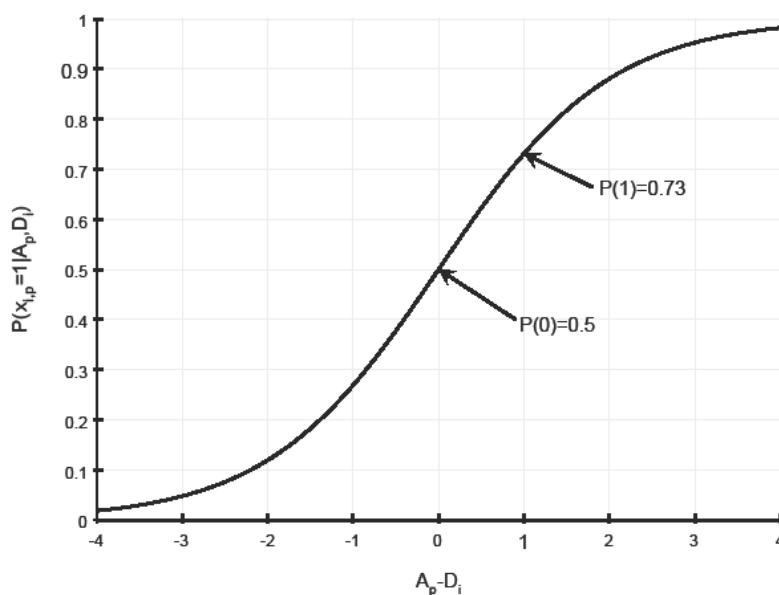


FIG. 1. The shape of the Item Characteristic Curve.

The distance between a person's ability and an item's difficulty (along the horizontal axis) is a measure of how well-suited the item is for the person, and is defined because the parameters are both on the same scale. For instance, if a person and item are judged to have the same parameters (i.e.  $A_p - D_i = 0$ ), the probability that the person answers the item correctly is defined to be 0.5. The choice of scale for the parameters is arbitrary; all that is required is that for a fixed distance between a person and an item there should be a fixed probability of answering the item successfully. The Rasch Model uses a logit<sup>‡</sup> scale for  $A_p$  and  $D_i$ .

A more difficult item or a more able person's parameters are larger positive numbers, and easier items/less able persons have larger negative values. Desirable values of  $A_p - D_i$  lie between -3 and 3 logits, corresponding to success-probabilities of 0.05 and 0.95. What it means to assign a probability for an event that has already occurred (the person has already answered the item), is discussed by Fischer and Molenaar (1995, §1.2).

### 2.3. Rasch Analysis

Given a test that produces dichotomous data (answers are either correct or incorrect), Rasch Analysis is performed in two stages:

1. Estimation of  $A_p$  and  $D_i$  for each person and item;
2. Test fit of model.

In the first stage, persons with zero or all items correctly answered are removed from analysis. This is because their ability parameters cannot be estimated by the test; it would be incorrect to say that someone scoring 100% had an ability of  $\infty$ , for one can never claim that the test contains the hardest question that it is possible to write. Similarly items on which

<sup>‡</sup> A logit scale takes a success-probability  $\pi$ , and transforms it to the scale  $\pi \rightarrow \log_e \left( \frac{\pi}{1-\pi} \right)$ .

every person was (un)successful are removed. Such persons and items are re-introduced after the other persons and items have had parameters estimated.

After this removal, we have a matrix of 0 and 1s, as in the left side of Fig. 2. Rasch Analysis produces estimates of the  $A_p$  and  $D_i$ , and then uses equation (1.1) to produce another matrix of success-probabilities, as shown on the right of Fig. 2. There are a number of statistical procedures that can be used to estimate the parameters and in our analysis we used the Joint Maximum Likelihood Estimation (JMLE), recently described by Baker and Kim (2004). This is a standard approach, and is used by the computer package Winsteps (Linacre, 2009).

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Person 1	1	0	1	1	1	1
Person 2	1	0	1	1	1	0
Person 3	1	1	1	1	1	1
Person 4	1	1	0	1	1	1
Person 5	1	1	1	1	1	0
Person 6	1	0	1	1	0	1

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Person 1	0.56	0.27	0.96	0.73	0.98	0.84
Person 2	0.59	0.48	0.88	0.75	0.65	0.12
Person 3	0.96	0.80	0.69	0.52	0.69	0.84
Person 4	0.68	0.77	0.58	0.99	0.77	0.68
Person 5	0.55	0.77	0.65	0.72	0.87	0.28
Person 6	0.73	0.29	0.85	0.80	0.42	0.54

FIG. 2. Illustrative matrices of corresponding observed data (left) and estimated probabilities (right). The remainder of the figures and tables in the document are those from our data, which is introduced in Section 3.

### 3. Research instrument and data

#### 3.1. Participants

The data that forms the basis of the Rasch Analysis undertaken in this study was collected from 164 first-year mathematics undergraduates at a central UK university. These undergraduates all had achieved grade A at A-level mathematics (or equivalent) and were midway through a course on basic real analysis given in the first term. When the task was administered the students had covered the topics of inequalities, sequences and series, and has been exposed to definitions of properties connected with these topics, presented both in standard formal statements and informal rephrasing of these statements.

In our study the students were given an example generation task (described below in further detail) during small-class tutorials, each involving approximately 25 students. They were given no assistance other than the information contained on the task sheets and none indicated they had insufficient time to complete the task. In the remainder of the report we shall refer to the example generation task as 'the task,' leaving the word 'test' for more general instruments.

#### 3.2. The task sheet and definition sheet

The task sheet asked student to give examples of (real) sequences that satisfied a combination of properties, or to state that the combination was impossible. There were 11 questions in total, presented in the order found in Table 1, below. We note that only one combination is impossible, question 10. While they were completing the task students had

access to a sheet which contained the formal definitions of the properties specified in the questions, but not the negations of these properties.

Tests such as the task given in this study have been used as a “window into the understanding of mathematics” of participants (Zazkis & Leikin, 2007). In particular, here they are used as a way of investigating students’ example spaces and identifying students’ conceptions and assumptions relative to the formal theory.

### 3.3. Coding of Answers

Answers were initially coded as correct or incorrect. More detailed breakdowns of the types of answer given for both correct and incorrect categories have been made subsequently, but have not been used in this analysis, because we are using the dichotomous Rasch Model. Data were entered into the model without distinguishing between different tutorial groups.

Answers were marked leniently with regard to minor misuses of notation (such as using the wrong type of brackets), but harshly when students gave general answers which implicitly included both correct and incorrect sequences. For instance, if a student gave the answer

$$a_n = ke^n, k \in \mathbb{R}$$

as an example of a ‘sequence that tends to infinity,’ the answer was marked incorrect; the sequence does not tend to positive infinity for negative values of  $k$ .

## 4. Results

In this section we first present the estimations of the difficulty parameters  $D_i$  for each question on our task. We then discuss the interpretation of these parameter estimates both generally and in terms of our task. In section 4.2 we then turn our attention to the students who sat the task, and explore responses that were ‘uncharacteristic’.

### 4.1. Preliminary Rasch Analysis – Difficulty of Items

The computer package Winsteps (Linacre, 2009) was used to analyse the data described in Section 3. The estimates of the difficulty parameters for each question can be found in brackets after each question (recall that negative numbers indicate easier questions).

1.	A strictly increasing sequence	(-1.83)
2.	An increasing sequence that is not strictly increasing	(0.53)
3.	A sequence that is both increasing and decreasing	(0.90)
4.	A sequence that is neither increasing nor decreasing	(0.66)
5.	A sequence that has no upper bound	(-4.27)
6.	A sequence that has neither an upper bound nor a lower bound	(0.26)
7.	A bounded, monotonic sequence	(-0.05)
8.	A sequence that tends to (positive) infinity	(-2.80)
9.	A sequence that tends to (positive) infinity that is not increasing	(3.93)
10.	A sequence that tends to (positive) infinity that is not bounded below	(1.13)
11.	A strictly increasing sequence that does not tend to (positive) infinity	(1.55)

TABLE 1. Questions asked in the sequence-generation task (difficulty parameter estimations in brackets).

This information is visually represented in Fig. 3., below.



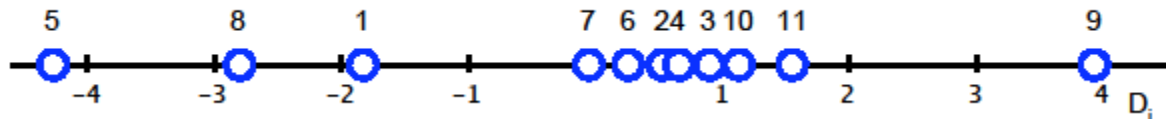


FIG. 3. Difficulty scale of the items

Interpreting such information will depend on the goal of the analysis. Test designers test might like to ensure that the questions are suitable for as wider population as possible, and so would require their items to cover a wide region (say  $-4$  to  $4$  logits), with questions spaced out along the interval. Such a test would be good at estimating the ability of persons with a wide variety of abilities, but would not have as much accuracy per person as a test designed for a specific ability group, where questions would be chosen from a narrower region. Such a targeted test would have good accuracy for the intended audience (i.e. those persons with a similar ability to the targeted questions), but it could only tell us a small amount about the persons whose ability was much higher or lower than the targeted region.

The questions in our task ranged in ability from very easy questions (5, 8 and 1), to a difficult question (9), but with the majority of questions in the interval  $(-0.05, 1.55)$ . This was desirable for a number of reasons. Firstly, because we had designed a research task and not an assessment test, some 'easy' questions were included to help put the students at ease (this is not to say that such questions are undesirable in a test). Secondly, if we were to find answers that were uncharacteristic (see the next section), it was important that most students felt that they could attempt the majority of questions, otherwise they might leave many questions blank. Of the students sitting our task, 60% were estimated to have ability parameters between 1 and 4 logits, and so the questions were indeed targeted so that most students were able to attempt them in principle.

It is outside the scope of this paper to analyse the estimates of the  $D_i$  parameters in any greater detail, but we note that question 9 was more than twice as difficult as the next most difficult question, 11 ( $D_9 = 3.93$ ,  $D_{11} = 1.95$ ). Both of these questions are about the relationship between sequences which tend to infinity and those which increase, an issue that will be revisited in section 4.3.

#### 4.2. Using Rasch Analysis to identify uncharacteristic responses within our data

In Section 2.3 we outlined the two stages of Rasch Analysis: estimating parameters and testing degree of fit. In Section 4.1 we explored the output of the first stage for the estimates of difficulty parameter, and in this section we will now consider the second stage, but with the person ability parameters instead.

Testing the degree of fit with the model involves comparing the observed responses with those estimated by the Rasch Model. Responses where there is a discrepancy between the observed and estimated values are said to misfit the model, and it is these misfitting responses that we call uncharacteristic, at least as far as the Rasch Model is concerned. We will argue that from both the position of a teacher and a researcher it is useful to identify, on a case-by-case basis, whether such answers are uncharacteristic in a wider (qualitative) sense of the word. If an able student gets an easy question incorrect, was it a slip of notation, or is there a fundamental problem with the student's view of the topic that only one question on the task brings into focus? Similarly, if a less able student only answers one

question correctly, but that question is one of the most difficult on the test, one can ask: might the student have guessed, or do the student's other answers indicate that that the question answered correctly is qualitatively different to the others? In the remainder of this paper we first consider the most uncharacteristic strings of responses, concerning ourselves with the students whose answers overall were uncharacteristic. Then, in Section 4.3, we change our focus to examine the most uncharacteristic individual responses.

The output in Table 2 lists those students whose responses most misfit the Rasch Model, in other words those students whose answers differ (in a chi-squared sense) from the estimations given by the model. Columns of interest are:

<b>Person</b>	Each row corresponds to one person
<b>Measure</b>	This is the estimate of a person's ability parameter
<b>Outfit - ZSTD</b>	This is the sum of squared standardized residuals, given as a Z-statistic. An Outfit-ZSTD value of $x$ corresponds to $100(x - 1)\%$ more variation between the estimated and observed values than if the model described that student's responses perfectly (Bond & Fox, 2007, p.239).

ENTRY	TOTAL			MODEL	INFIT	OUTFIT	IPT-MEASURE	EXACT MATCH					
NUMBER	SCORE	COUNT	MEASURE	S.E.	IMNSQ	ZSTD	IMNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Person
141	10	11	3.56	1.25	2.10	1.3	9.90	4.1	A-.27	.39	81.8	92.2	141
147	1	11	-3.92	1.30	2.32	1.6	8.95	2.8	B-.10	.49	81.8	92.3	147
148	7	11	1.03	.76	1.17	.7	4.50	1.9	C .33	.53	90.9	74.3	148
142	7	11	1.03	.76	2.12	3.3	3.82	1.7	D .01	.53	36.4	74.3	142
47	6	11	.47	.75	1.49	1.7	3.37	1.7	E .30	.57	54.5	73.4	047
52	7	11	1.03	.76	1.77	2.4	2.45	1.2	F .21	.53	54.5	74.3	052

TABLE 2. Output of Rasch Analysis, ordered by greatest Outfit-ZSTD statistic

Acceptable values for the misfit statistics are for those routinely accepted for a  $t$  statistic, i.e. -2 to 2 (Bond & Fox, 2007, p.57). Looking at the row which describes person 52, for instance, we see that the Outfit-ZSTD statistic is 1.2, which corresponds to 20% more variation than if the model described the student's responses perfectly. This is an indicator that my task was suitable for the students who sat it; only two students' Outfit-ZSTD statistic was greater than 2 (and no student had a value of less than -2).

We now analyse sections of this table in more detail. We first note that the students that most misfit the model have a wide range of ability parameter estimates; values in the 'measure' column of Table 2 vary considerably. There are students who scored highly on the test, such as student 141 who answered 10/11 questions correctly, and those who scored badly, such as student 147 who answered 1/11 questions, together with a spread of values between these.

The large Outfit-ZSTD values for students 141 and 147 indicate that it is likely that student 141 answered an 'easy' question incorrectly, and that student 147 answered a 'tough' question correctly and the remainder incorrectly. This is indeed the case: student 141 was incorrect on question 1 ( $D_1 = -1.83$ ) because the answer was given in the form of a recursive sequence, but without a starting value. No other sequence on this student's sheet

was presented recursively. Student 147 wrote mathematical nonsense, such as  $\pm\sqrt{a_n}$  for all but one question, giving a correct answer  $a_n = 1$  for question 4 ( $D_4 = 0.66$ ).

#### 4.3. A more detailed look at two uncharacteristic answers

Whilst the data given in Table 2 gives the degree of misfit per person, it can also be useful to look at the residuals on a cell-by-cell basis. This means we look not at which person gives the least characteristic set of responses, but which individual responses were most uncharacteristic. These first few such responses are listed in Table 3. Comparing the observed and estimated (expected) columns, we see that most differ by a large amount (the maximum theoretical difference is  $\pm 1$ ).

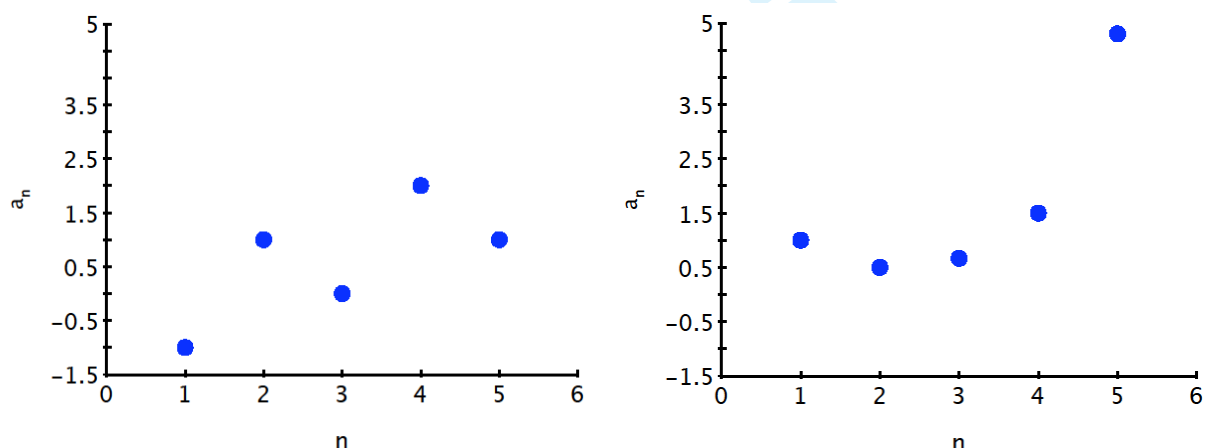
OBSERVED	EXPECTED	RESIDUAL	ST. RES.	MEASDIFF	Item	Person	Item	Person
0	1.00	-1.00	-14.59	5.36	1	141	I0001	141
1	.01	.99	9.83	-4.57	4	147	I0004	147
0	.98	-.98	-6.69	3.80	8	148	I0008	148
1	.03	.97	5.44	-3.39	9	47	I0009	047
0	.95	-.95	-4.59	3.05	2	161	I0002	161
0	.95	-.95	-4.59	3.05	2	158	I0002	158

TABLE 3. The output of misfitting individual responses

We have already discussed person 141's response to question 1 and person 147's response to question 4 in the last section. Analysis of person 148 and 47 is similar, but it is interesting that after these four, the next two uncharacteristic responses (from person 161 and 158) were both for the same question. The students were estimated to answer the question correctly with a probability of 95%, and yet both students answered this question incorrectly. We now look at their responses to this question in more detail.

**Question 2.** [Give an example of] an increasing sequence that is not strictly increasing

The responses to this question by student 161 and 158 were as follows. We have graphed the sequences for clarity (neither student sketched a graph as part of their answer).



$$\{a_{2n}\} = n \quad \{a_{2n-1}\} = n - 2$$

$$a_n = \frac{n!}{n^2}$$

Figure 4: Responses to question 2 from student 161 (left) and student 158 (right). Answers are presented in the same format as given by the students.

1  
2  
3 Both these sequences are consistent with the same underlying misconception relative to the  
4 formal theory. Both contain terms which are less than their predecessors, and so according  
5 to the formal theory are not increasing. However it is easy to see why such students may  
6 give answers similar to those in Fig. 4., because in an 'everyday' sense of the word their  
7 sequences are increasing; the terms are generally going up overall. Such misconceptions  
8 linked to the 'everyday' meaning of words have been called spontaneous conceptions  
9 (Cornu, 1991). Whilst we are not claiming that a Rasch Analysis will always identify the  
10 same qualitative features in responses identified in this manner, it is clear that answers that  
11 are identified may be worthy of closer examination. This is especially important when the  
12 volume of data prevents a detailed analysis of each individual response.  
13  
14  
15  
16  
17

## 18 5. Conclusion

19 In this paper we have introduced the Rasch Model, and demonstrated two applications of  
20 Rasch Analysis using the results of an example-generation task given to mathematics  
21 undergraduates. We first ranked the questions of our task and discussed the implications of  
22 this ranking, both in terms of designing tests and analysing tasks. We then looked at those  
23 responses which were uncharacteristic to the Rasch Model, and argued that such responses  
24 are worthy of further examination.  
25  
26  
27

28 The use of Rasch Analysis has allowed us to easily and quickly identify uncharacteristic  
29 responses within a large data set. This would have been difficult, or not impossible, using a  
30 script-by-script individual analysis. By examining the responses which have emerged from  
31 the Rasch Analysis as uncharacteristic, we have identified types of answer that were  
32 unexpected statistically, and by qualitatively examining these scripts we have identified  
33 mathematical misconceptions relative to the formal theory. This approach could be useful  
34 not just for researchers in mathematics education, but also for lecturers of mathematics  
35 wishing to identify topics of difficulty in their class assignments.  
36  
37  
38

## 39 REFERENCES

- 40 BAKER, F. B. & KIM, S.-H. (2004) *Item response theory: Parameter estimation techniques*  
41 (second ed.). New York: Marcel Dekker.  
42  
43 BOND, T. G. & FOX, C. M. (2007) *Applying the Rasch model*. Mahwah, New Jersey:  
44 Lawrence Erlbaum Associates.  
45  
46  
47 BRADLEY, K., SAMPSON, S., & ROYAL, K. (2006) Applying the Rasch rating scale model to  
48 gain insights into students' conceptualisation of quality mathematics instruction,  
49 *Mathematics Education Research Journal*, **18**, 11–26.  
50  
51  
52 CALLINGHAM, R., & BOND, T. (2006) Editorial: Research in mathematics education and Rasch  
53 measurement, *Mathematics Education Research Journal*, **18**, 1–10.  
54  
55  
56 CHEN, C.-M., LEE, H.-M., & CHEN, Y.-H. (2005) Personalized e-learning system using item  
57 response theory, *Computers & Education*, **44**, 237–255.  
58  
59 CORNU, B. (1991). Limits. *Advanced mathematical thinking* (D. Tall ed.). Netherlands: Kluwer  
60 Academic Publishers, pp. 153–166.

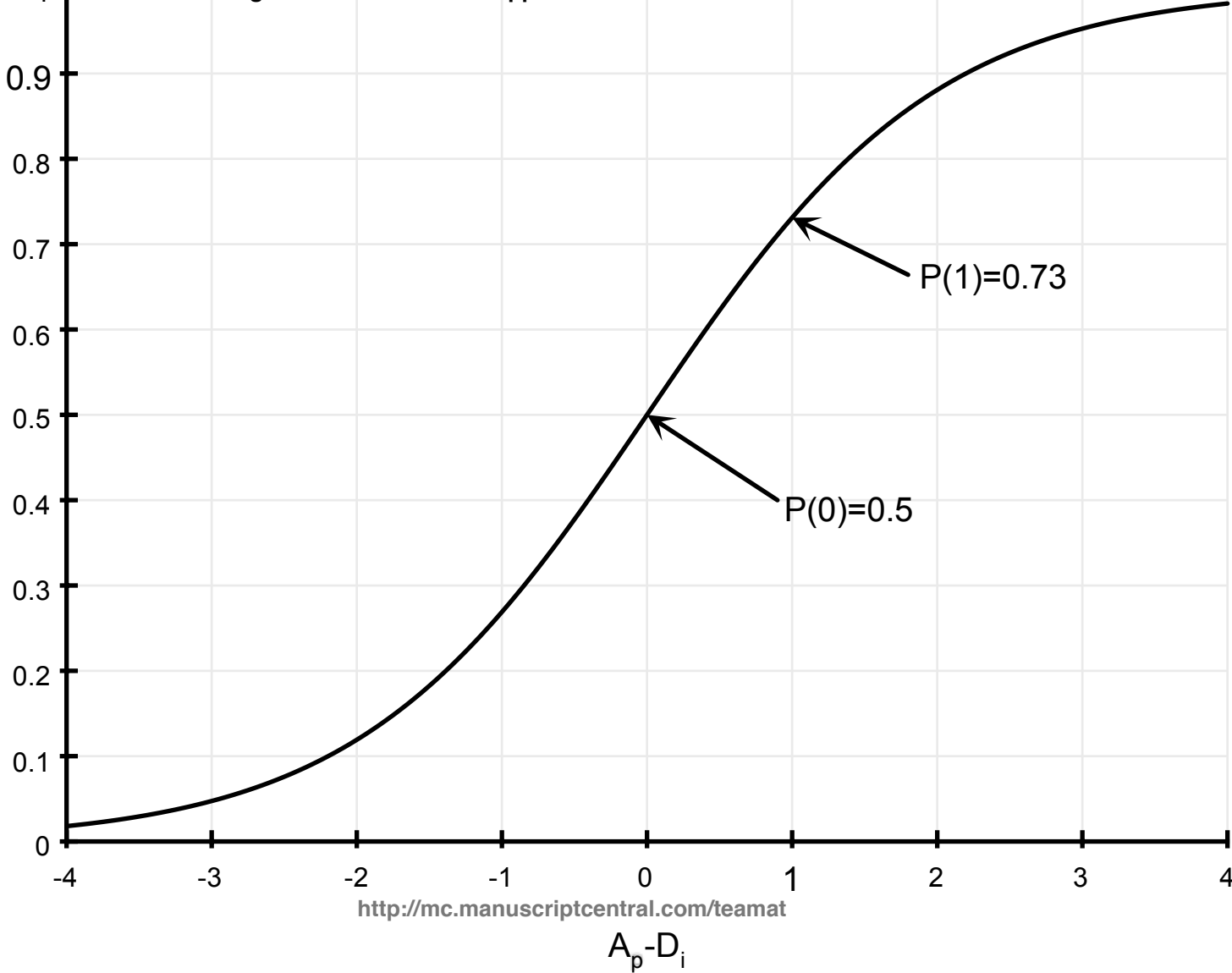
- 1  
2  
3 DOBBY, J., & DUCKWORTH, D. (1979). Objective assessment by means of item banking.  
4 London: Evans Bros for the Schools Council / Methuen.  
5  
6  
7 EMONS, W., SIJTSMA, K., & MEIJER, R. (2005) Global, local, and graphical person-fit analysis  
8 using person-response functions. *Psychological Methods*, **10**, 101–119.  
9  
10 FISCHER, G. & MOLENAAR, I. (1995) *Rasch models: foundations, recent developments and*  
11 *applications*. New York: Springer-Verlag.  
12  
13  
14 GRIMBEEK, P., & NISBET, S. (2006) Surveying primary teachers about compulsory numeracy  
15 testing: combining factor analysis with Rasch analysis, *Mathematics Education*  
16 *Research Journal*, **18**, 27–39.  
17  
18  
19 HAINES, C., & CROUCH, R. (2001) Recognizing constructs within mathematical modelling,  
20 *Teaching Mathematics and its Applications*, **20**, 129–138.  
21  
22 KLINE, T. (2005) *Psychological testing: A practical approach to design and evaluation*. Sage  
23 publications.  
24  
25  
26 LINACRE, J. (2009) *Winsteps* (version 3.68.2) [computer software]. Beaverton, Oregon:  
27 winsteps.com.  
28  
29  
30 RASCH, G. (1980) *Probabilistic models for some intelligence and attainment tests* (expanded  
31 ed.). The University of Chicago Press.  
32  
33 REISE, S. (2000) Using multilevel logistic regression to evaluate person-fit in IRT models.  
34 *Multivariate Behavioral Research*, **35**, 543–568.  
35  
36 RYAN, J., & MCCRAE, B. (2006) Assessing pre-service teachers' mathematics subject  
37 knowledge. *Mathematics teacher education and development*, **7**, 72–89.  
38  
39  
40 STACEY, K., & STEINLE, V. (2006) A case of the inapplicability of the Rasch model: mapping  
41 conceptual learning, *Mathematics Education Research Journal*, **18**, 77–92.  
42  
43  
44 WATSON, J., KELLY, B., & IZARD, J. (2006) A longitudinal study of student understanding of  
45 chance and data, *Mathematics Education Research Journal*, **18**, 40–55.  
46  
47  
48 WAUGH, R. F. (1998) The course experience questionnaire: a Rasch measurement model  
49 analysis, *Higher Education Research & Development*, **17**, 45–64.  
50  
51  
52 ZAZKIS, R. & LEIKIN, R. (2007) Generating examples: from pedagogical tool to a research  
53 tool, *For the Learning of Mathematics*, **27**, 15-21.

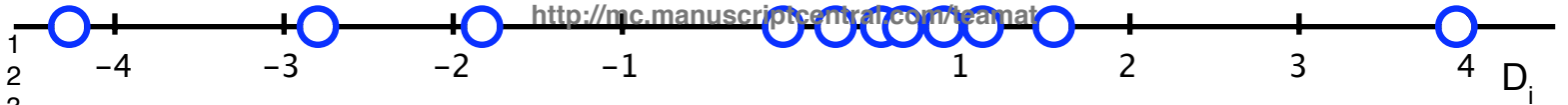
54 **Antony Edwards** is a PhD research student at Loughborough University's Mathematics  
55 Education Centre. He is investigating how example generation tasks can be used to explore  
56 students' conceptions of real-sequences, primarily using Marton and Booth's methodology of  
57 Phenomenography. He is also interested in combining qualitative and quantitative  
58 approaches to analysing data from such example generation tasks.  
59  
60

1  
2  
3 **Lara Alcock** teaches mathematics at Loughborough University's Mathematics Education  
4 Centre, and has also taught both mathematics and mathematics education at other  
5 universities in the UK and USA. Her research focuses on students' understanding of real  
6 analysis, the role of examples in mathematical thinking and proof comprehension. She also  
7 designs pedagogical tools for helping students learn mathematics and resources to assist  
8 lecturers in understanding how their students learn mathematics.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36







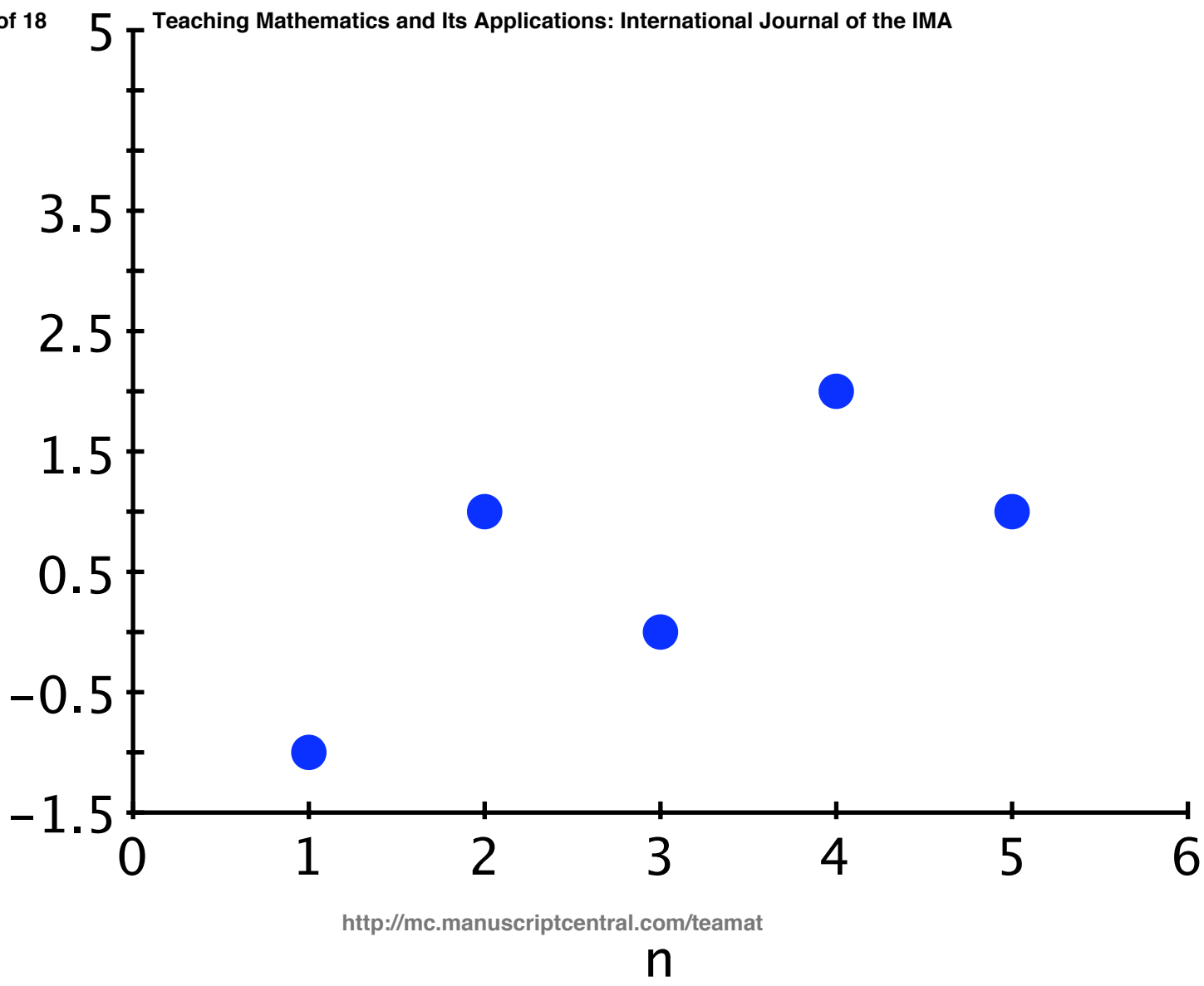
ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT IMNSQ	OUTFIT ZSTDIMNSQ	PT-MEASURE ZSTDICORR.	EXACT EXP.	MATCHI OBS%	Person EXP%				
1														
2	141	10	11	3.56	1.25	2.10	1.3	9.90	4.1	A-.27	.39	81.8	92.2	141
3	147	1	11	-3.92	1.30	2.32	1.6	8.95	2.8	B-.10	.49	81.8	92.3	147
4	148	7	11	1.03	.76	1.17	.7	4.50	1.9	C .33	.53	90.9	74.3	148
5	142	7	11	1.03	.76	2.12	3.3	3.82	1.7	D .01	.53	36.4	74.3	142
6	47	6	11	.47	.75	1.49	1.7	3.37	1.7	E .30	.57	54.5	73.4	047
7	52	7	11	1.03	.76	1.77	2.4	2.45	1.2	F .21	.53	54.5	74.3	052

6  
5  
4  
3  
2  
1

DATA	OBSERVED	EXPECTED	RESIDUAL	ABS. RES.	MEASUREMENT	Person	Item	Page
0	0	1.00	-1.00	14.59	5.36	1	I0001	141
1	1	.01	-.99	9.83	4.57	4	I0004	147
0	0	.98	-.98	6.69	3.80	8	I0008	148
1	1	.03	-.97	5.44	3.59	9	I0009	047
0	0	.95	-.95	4.59	3.05	2	I0002	161
0	0	.95	-.95	4.59	3.05	2	I0002	158

<http://mc.manuscriptcentral.com/teamat>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

