# Language, Contextual and Cultural Constraints on Examination Performance

Alastair Pollitt, Carolyn Marriott & Ayesha Ahmed

University of Cambridge

**A paper presented to the**

**International Association for Educational Assessment,**

**in Jerusalem, Israel, May 2000.**

The opinions expressed in this paper are those of the authors and should not be taken as official policy of the University of Cambridge Local Examinations Syndicate or any of its subsidiaries.

**Contact details**:

Research and Evaluation Division

University of Cambridge Local Examinations Syndicate

1 Hills Road

Cambridge

UK  CB1 2EU

Telephone: +44 1223 553847

Fax: +44 1223 552700

e-mail: pollitt.a@ucles.org.uk

# Language, contextual and cultural constraints on examination performance.

## Summary

*In a multicultural world educational assessment very often means that many students are assessed in a language which is not their mother tongue and in the company of other students whose culture they do not fully share. What special difficulties does this raise for these students?*

*The paper is based on a case study of a student recently immigrated to England from Pakistan who seemed to be under-performing in mathematics. Rasch analysis of his responses to two mathematical tests highlighted certain questions that he got 'surprisingly' right and others, more interestingly, that he got 'surprisingly' wrong. Linguistic analysis of the questions and post-test interviews with the student were used in order to understand the answers that he gave.*

*Clear conclusions are drawn of three kinds:*

*Linguistic:*

> *The uses of 'ordinary' English words with special meanings can cause unexpected problems for L2 students;*

*Contextual:*

> *'Real world' contexts can so complicate the task, if the context is not familiar, that comprehension and task solution are prevented;*

*Cultural:*

> *Language and context may interact in subtle ways such that apparently easy questions become impossible for culturally disadvantaged students.*

*The paper interprets the student's responses to ten different mathematics questions; these questions and his responses to them are shown in the appendix.*

## Introduction

Educational assessment has never been simple and homogeneous. A nation may contain several cultures within its borders, or receive a substantial influx of immigrants. Increasingly, educational examinations may become trans-national like other service industries. In such circumstances it becomes more important than ever to understand how students whose language or culture differs from those of the largest group may be handicapped.

This paper is based on a case study. A 16 year old Pakistani student arrived in Britain in March; to give him time to adjust to life in the UK he was placed in a Year 10 class, with students generally about a year younger than him who were due to take the school leaving (GCSE) examinations 14 months later.

In September he was placed in a *foundation* class, but soon expressed concern that the work was "too easy"; teachers, however, were reluctant to promote him to *intermediate* as long as his English was too weak to let him participate fully in normal classes. He continued to receive special language help, alongside a few other recent arrivals.

In the following January he attempted the foundation, intermediate and higher papers in the 'mock' GCSE examination, which were marked internally by the school. This paper is concerned with his performance on the first two of those papers.

## Data

The original case study dealt with MS in detail, and included the following sources of evidence:

1 Discussions with MS's teachers.

2 Comparison of the mathematics syllabuses in Pakistan and England.

3 Analysis of linguistic characteristics of the questions and papers.

4 Question level data from 38 other students who took the foundation and intermediate papers.

5 Written responses from MS to all of the questions in these three papers.

6 Delayed post hoc interviews with MS about his attempts to answer both papers.

## Preliminary analysis

Concern about MS was initiated when, in late September, he himself complained that the work in his foundation level course was "too easy". His support teacher suggested that he might be promoted to the intermediate level, but his mathematics teacher was reluctant to move him so long as his English was too poor for him to participate fully in the class. Nevertheless, MS's teachers agreed that he showed signs of good mathematical ability. The school's decision was that he should be left in the foundation group at least until the 'mock' GCSE examination in January.

A cursory analysis of the syllabus he had followed in Pakistan was enough to show that he had learned a much more formal version of mathematics than is normal in the UK nowadays. The Pakistan syllabus emphasised sets and functions as the fundamental concepts of mathematics rather than focusing on applications of mathematical principles to problems in the real world, or indeed on the recognition of mathematical principles in real contexts.

### Rasch analysis

MS made many errors in his two papers, but then so did most other students. In order to decide which errors were worth investigating further, a Rasch analysis of a full class data set was carried out. The school supplied a sample of 39 students' responses to the two papers, which were enough to create a 'fixed' background against which to evaluate MS's performance.

There are two key statistics considered in this procedure:

(a) the unweighted misfit statistic, and

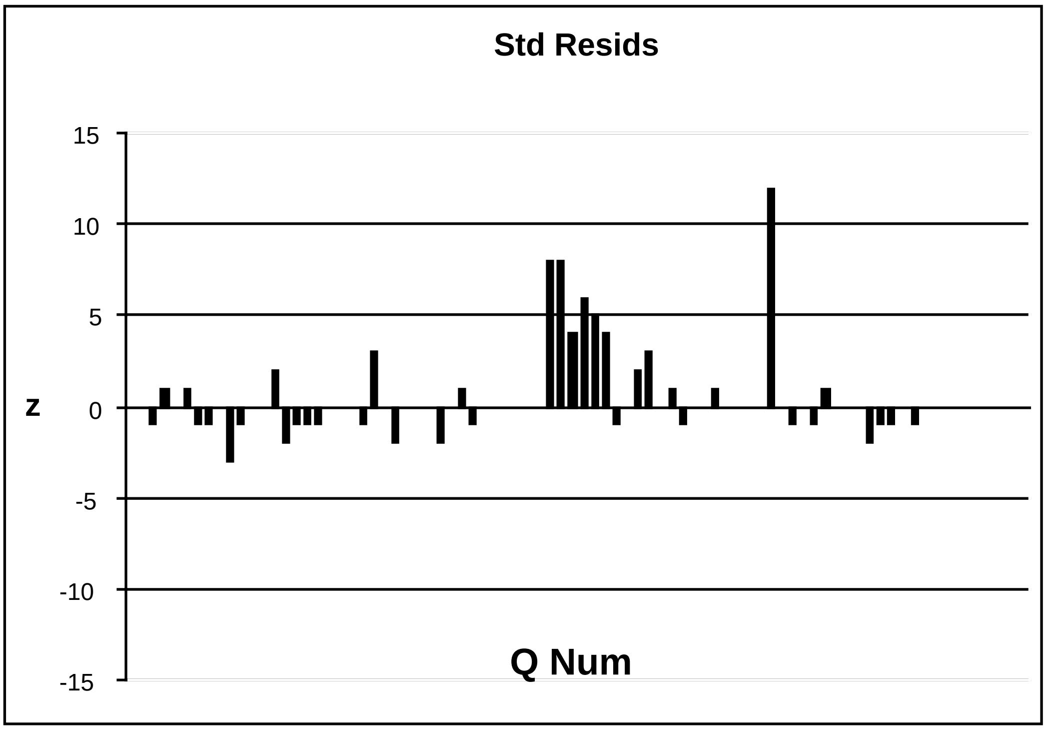
(b) the standardised residual for each response.

It is easier to consider these in reverse order. Once an overall measure of a student's ability has been estimated from the total score, together with an overall estimate of each question's difficulty[§] , it is possible to predict how each student 'should' perform on each question. The difference between the observed score on the question and this prediction is the *residual*, and this is normally standardised by dividing it by its own standard error (see Wright & Masters, 1982, for details of this procedure). There is therefore one standardised residual for the response made by each student to each question, and these are best interpreted as measures of how surprising each response is, given our overall estimates of the student's ability and the question's difficulty. Note that the standardised residual will always be negative when the student does less well than predicted, and positive when the student does better than expected.

The unweighted misfit statistic is defined as the mean square of these standardised residuals, summed either across each question (item analysis) or across each student (person misfit analysis). Often (as here) this statistic is converted to a standardised normal, or z, form for ease of interpretation. The interest here is in the misfit statistics of students, and as with 't' or 'z' statistics any value greater than ±2 may be taken to indicate a student whose pattern of performance across the various questions is unusual enough to merit further study.

From the 39 students for whom we had data, four (including MS) gave overall misfit statistics greater than ±2, but only MS registered more than 4. If the data and the model had fitted perfectly we would expect two misfits greater than 2, and so the overall fit, apart from MS, was reasonable. MS's value of 4.37 is unusual enough to arise by chance with a probability of less than 0.001%.

Inspection of MS's standardised question residuals gives the picture below.




There are several residuals of around 2 or 3, but the most notable feature is the seven residuals greater than +4. These are all questions where MS scored very well, indicating an ability much higher than his overall score of 43 out of 127 would lead us to expect.

---

[§] If a question is worth more than one mark there is one 'difficulty' for each mark; this does not alter the procedure for calculating the two statistics.

These seven responses belong to three questions on the intermediate level paper, P2; the questions and MS's responses are shown in Appendix A:

**P2Q5a, b** **(M)**

There are two parts, requiring numbers from the text to be inserted into a given algebraic equation, so that the result can be calculated. Only one other student in the sample of 39 was able to answer both correctly.

**P2Q6a, b, c, d** **(M)**

This is similar, except that it is not necessary to extract numbers from the text; a diagram is given with the numbers and the required angles - a, b, c and d - are marked on it. It was not necessary for MS to understand the context, or even to know what the "leaning tower of Pisa' is in order to see the point of the question.

**P2Q12b** **(mV)**

This is the most remarkable: MS was the only student in the whole group of thirty-nine able to answer it correctly. Yet, if we are to believe his overall score accurately indicates his ability, then the odds against him doing so were 163,000 to 1. He failed on part (a), and in interview it seemed that he was not sure what the question was asking; he guessed, and guessed correctly, what part (b) demanded and, given this good fortune, was able to carry out the calculation correctly even though none of his peers could do so.

MS's right answers to these difficult questions are clear evidence that he was indeed quite able at mathematics, at least in some conditions.

Given the evidence of his potential performance, the real surprise is that he got more than half the questions in the examination wrong. In the remainder of the paper we shall look at the most surprising of these wrong answers, all of those where the original standardised residuals were at all negative. In most cases we can see, either from what he wrote or from what he said at interview, that there were specific linguistic or cultural reasons why he had difficulty.

## Linguistic analysis

Mathematics is often considered as the study of number, magnitude, shape and the symbolic representations of these elements and their relationships. These form the lexicon and the syntax of a "universal language" of mathematics: but is it truly a language? Pimm (1991) argues that it is not, since "there is no group of people for whom mathematics is their first language". Others (eg Austin and Howson, 1979) see maths as an activity that employs many languages, from the vernacular to the language of formal logic, or as an activity using a hybrid language (eg Kane, 1970). A compromise between these various views and the notion that maths is in fact a language in its own right would be to consider mathematical language as a distinct *register*, that is, " a set of meanings that is appropriate to a particular function of language, together with the words and structures which express these meanings" (Halliday, 1978). The important feature of a register is that it represents a set of conventions that lie behind the simple lexical and syntactic surface of the language, and that must be recognised by the reader as a critical part of the context of the text.

### Mathematical English and Vernacular English

In the original study on which this paper is based, Marriott (1993), a four category scale was created to classify questions according to how great a demand they made on students.

| MATHEMATICAL | MATH/vern | math/VERN | VERNACULAR |
|---|---|---|---|
| **M** | **Mv** | **mV** | **V** |

<----------------------------------------------------------------------------------------------->

Category M questions presented the problem clearly in mathematical symbols or in simple language using the names of mathematical units, shapes or operations; the essential point is that students could understand what to do by understanding *only* mathematical symbols and formal mathematical English.

Questions in category Mv can only be answered correctly by students who understand some vernacular English lexical items, but these are words for which specifically mathematical meanings are learned in the classroom; examples include *total*, *bar chart*, *area*, *scale* or *plot*.

Category mV contains those where the problem is presented in vernacular English, but where the language gives strong clues to the kind of mathematics needed. Examples include *altogether*, which strongly cues addition, or *left over*, which equally strongly indicates subtraction.

Finally, category V contains questions which express a mathematical problem but which contain no lexical cues that may have been learned in the maths classroom; the candidate must understand the language well enough to see through the vernacular English to the problem.

We have already seen three questions where MS did 'surprisingly well'. They are classified as:

| | | | | |
|---|---|---|---|---|
| P2Q5 | M | | | |
| P2Q6 | M | | | |
| P2Q12b | | | mV | |

In the rest of the paper we will look at the questions where he did 'surprisingly badly'. They are classified as:

| | | | | |
|---|---|---|---|---|
| P1Q10a | | | | V |
| P2Q7a | | Mv | | |
| P2Q10aii | | | mV | |
| P1Q17 | | | | V |
| P2Q2a | | | mV | |
| P1Q2b | | | mV | |
| P1Q13b | | | | V |

The diagnosis of MS's difficulties with mathematics seems obvious:

***MS could handle mathematics well, but had serious difficulties with problems that were expressed in vernacular English.***

## Analysis of question responses

We now turn to a qualitative analysis of all of the 'surprising' wrong answers that MS gave, or the questions where he did much less well than we would expect. The questions are considered in three sets, according to whether they seem to indicate difficulties of a linguistic, contextual or cultural kind:

linguistic:

- P1Q10a
- P2Q7a
- P2Q10aii

contextual:

- P1Q17
- P2Q2a

cultural:

- P1Q2b
- P1Q13b

The questions and MS's response are given in Appendix B.

### *Language*

**P1Q10a** **(V)**

In this question we see that MS was unable to resolve a complex anaphoric reference ". . . then divide *that* . . .". The appropriate referent is the implicit *result* of subtracting of the age from 34, rather than either of the possible explicit noun phrases "the age" or "34". MS has chosen the simplest possible referent, "the age". His answer to part (b) shows that he misunderstood the meaning of *work* in this sentence, taking it to signify 'labour' rather than 'function'. He therefore failed to meet the objective the question was designed to assess:

> *understand a simple formula expressed in words and substitute numbers into it to calculate the value of the subject of the formula.*

Given the diagnosis above, this should come as no surprise.

**P2Q7a** **(Mv)**

Despite this being coded Mv, that is using specifically mathematical language, MS failed to understand the task. In interview he reported that he did not know the term *pie chart*, and the interviewer inferred that he was misled by the phrase "on the packet"; he concluded that he was to draw the information *on the packet*.

**P2Q10aii** **(V)**

MS's answer of "6cm" for part (a) was correct. In part (b) he measured the height of the 'wall' part of the drawing without the roof part, showing that he misunderstood the force of *actual*; he did not recognise it as an invitation to apply the "scale of 2cm to 1 m" to the result of part (i). *Actual* is not in any sense a mathematical term, and his English was not good enough to recognise that in this register it was meant to contrast the model with the real world.

### *Context*

When a student is suffering difficulties with language, it is hardly a surprise to find that he is unable to make use of a complex context to disambiguate what he reads. Two examples show this very clearly.

**P1Q17** **(V)**

The language problem here is subtle. in MS's first language (Urdu) there are two words that might be used to translate *height*. One refers to height from the ground, while the other indicates how tall something is; in English the same word fulfils both roles. The way the question was phrased *3.5 metres high* misled him to assume the first meaning; had it said *a 3.5 metre statue*, or t*he height was 3.5 metres*, he would have been led to the correct meaning. Instead, he assumed the statue was 3.5 metres above the ground and, quite reasonably. doubled it to get the height of the building.

**P2Q2a (mV)**

Here he seems to have understood exactly what the question required - except for one simple error. He actually calculated *two* answers, the number in 12 hours and the number in 24 hours. Which of these means "in a day"? In English we talk of a week being 7 *days* long, or of 365 *days* in a year, both of which correspond to the 24 hour meaning expected here; but we also contrast *day* with *night*, and refer to *day*light, implying that a day is only half of a complete revolution of the earth, or just 12 hours long. Which is the meaning here? As in the last question, MS's first language uses two different words: *din* signifies the 12 hour day, while *dinraath* (literally day-night) signifies the 24 hour period. When he was asked at interview why he had not chosen 24 as the number of hours in a day his response was "That's day and night".

## *Culture*

When we move from language itself into contexts that are culture specific the difficulties become even more demanding. In the final two questions we see MS struggling to make mathematical sense of problems that must have seemed very unfamiliar to him. In both cases the interview made it quite clear how he had failed to understand cultural aspects of the problem.

**P1Q2b (mV)**

Part (a) shows that MS had no problem reading the scale on the weighing machine, interpolating between the numbers and finding the correct value. But he then continued in the same vein, trying to interpolate between the charges in the table to find the 'best' cost for the parcel. He thought he should apply a good mathematical technique to part (b), just as he had in part (a). Unlucky MS; it is unlikely that a 16 year old, recently immigrated from Pakistan, had much experience of posting parcels in British post offices.

**P1Q13b (V)**

It is even more unlikely that MS had much experience of parking cars in British car parks. In interview, as well as in his answer to part (a) he showed that he interpreted "after lunch" in the context of school, where lunch came at 1 o'clock. Therefore she was parked for only about 15 minutes, since "her ticket was stamped 1315". But it was clear that MS did not really understand the meaning of this phrase, and so in part (b) he returned to the idea that she arrived just after 1 o'clock; then finding that she left at 5 past 7 in the evening he 'correctly' inferred that she had been parked for more than 6 hours. Of course, his failure to connect parts (a) and (b) means that his whole answer to the question was inconsistent, but the interview made it clear that he was, in fact, struggling to interpret the story throughout, and had no confidence in the truth of his response in either part. Yet he could clearly handle the mathematics and, in this case, was not tempted to try to interpolate or extrapolate from the values in the table.

## *Conclusions*

Mathematical symbols and mathematical English may be problematic for native speakers of English taking mathematics tests, but for students like MS, where English is a second language, the reverse is true. Vernacular English is, of course, a problem and we may be testing it more than mathematical achievement when we insist in contextualising all of our questions for these students.

In our description of the question answering process (Pollitt & Ahmed, 1999) the first stage in responding involves reading, or comprehending, the question. It is useful to consider carefully a sentence on reading comprehension by Philip Johnson-Laird:

> *words are cues to build a familiar mental model* (Johnson-Laird, 1983)

*Reading the question* involves building a mental model of the problem. In an examination question the words are intended to provoke specific schemas (Bartlett, 1932), where a schema is set of concepts, relationships and affects that are stereotypically associated in a person's memory. Students are expected to activate *appropriate* schemas and to build a particular mental model of the situation indicated by the words by combining and elaborating these schemas. Of course, each individual will build their own unique model, but it is expected that these will all share certain relevant features that will help with answering the question. The art of question writing is to control this process of model building so that the intended skills will be assessed in a way that is fair to all students.

It is important for question writers to remember that mental model building is not exclusively, or even principally, under the student's conscious control. Cue words provoke schemas, and it is only then, and with the schemas that have been provoked, that the student can deliberately seek or plan an answer. Evans (1989) describes what he calls *bias* in human reasoning in these terms:

> *... many biases are caused by preattentive or preconscious heuristic processes which determine selective encoding of psychologically "relevant" features of the problem.*

Concepts must be activated in the student's mind *before* they become available for conscious reasoning, and it is the function of the words, and diagrams, in the question to provoke appropriate concepts in the mind of any well prepared student.

Thus the model building process will go wrong if the student reacts inappropriately to the cue words. This may happen if they don't understand the language well enough, or if they are not familiar with the culture that formed the stereotypical schemas. These two kinds of difficulty are well known, to both students and their teachers. This research suggests, though, that a greater danger may lie *between* these two familiar sources of difficulty.

Recently immigrated students will be aware of the demands that vocabulary, grammar and culture will place on them. Neither they nor their examiners, however, are likely to recognise how misleading 'simple' words can be when they are used in a subtly different mathematical register. It is difficult even for expert language teachers to help students become more alert to the details of register, and dictionaries are unable to indicate all of the subtle changes in meaning that register may imply.

These differences in meaning will pose unfair difficulties for immigrant students even though, for us, they are as clear as day and night.

## References

**Austin, JL & Howson, AG**, 1979. Language and mathematical education. *Educational Studies in Mathematics*, **10**, 161-197.

**Bartlett, FC**, 1932. *Remembering*. Cambridge, Cambridge University Press.

**Evans, J St BT**, 1989. Bias in human reasoning. Lawrence Erlbaum Associates, Hove.

**Halliday, MAK**, 1978. Language as social semiotic: the social interpretation of language and meaning. Edward Arnold, London.

**Johnson-Laird, PN**, 1983. *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, Mass, Harvard University Press.

**Kane, RB**, 1970. The readability of mathematics textbooks revisited. **Mathematics Teacher**, **63**, 579-81.

**Marriott, JC**, 1993. *ESL in mathematics: why 'easy' us more difficult*. MPhil dissertation, Research Centre for English and Applied Linguistics, University of Cambridge,

**Pimm,D,** 1991. Communicating mathematically. In K Durkin & B Shire (eds) *Language in mathematical education: Research and Practice*, Open University Press, Milton Keynes.

**Pollitt A & Ahmed, A**, 1999. *A New Model of the Question Answering Process.* Paper presented to the International Association for Educational Assessment, Bled, Slovenia, May 1999.

**Wright, BD & Masters, GN**, 1982. *Rating Scale Analysis*, MESA Press, Chicago.

# Appendix A

**P2Q5a, b** **(M)**

**P2Q6a, b, c, d** **(M)**

**P2Q12b** **(mV)**

# Appendix B

## P1Q10a (V)

To work out how many hours of sleep young people need, a magazine suggested this rule.

"*SUBTRACT THE AGE FROM 34, THEN DIVIDE THAT BY 2.*"

(a) How many hours of sleep do these people need?

(i) Sarah, aged 16.

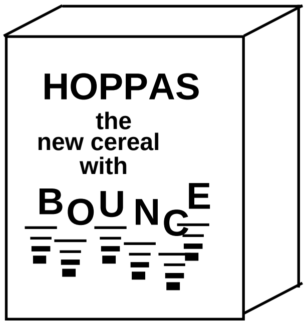
8 hours

…

(b) Why will the rule not work for all ages?

because the old people can't do work . They
has lost there energy and children are
????? so they can't do work with school .

P2Q7a (Mv)




A new breakfast cereal 'HOPPAS" gives this nutritional information on the packet.

| | |
|---|---|
| Protein | 10 % |
| Carbohydrate | 73 % |
| Fat | 2 % |
| Fibre | 7% |
| Vitamins/minerals | 8% |

(a) Draw a pie chart to show this information.

***P2Q10aii*** ***(V)***

(a) (i) On the side view, measure the line that shows the full height of the shed.

6 cm

(ii) How high is the actual shed?

4 cm

***P1Q17*** ***(V)***

This is a picture of the stage and rearwall of the Roman Theatre in Orange, France.

The statue of Augustus in the middle (not including the base) is 3.5 metres high.

Estimate the full height of the building.

3.5 + 3.5 = 7 m

..........................................................................................................................

***P2Q2a*** ***(mV)***

It is claimed that in Florida there are eleven lightning strikes every minute.

(a) How many is this in a day?

7920 in a day

in one minute = 11 light

in 12 hours = 720 x 11 = 7920

in 24 hours = 1440 x 11

= 15840 lights

***P1Q2b*** ***(mV)***

***postal charges***

***P1Q13b*** ***(V)***

***"lunch" and car park***