Contrasts between student and examiner perceptions of the nature of assessment tasks

D Clerk, D Naidoo and S Shrivastava

School of Physics, University of the Witwatersrand, Private Bag 3, WITS 2050

douglas.clerk@wits.ac.za

Abstract: For several years an on-going study has examined student performance in relation to the various *types* of assessment task encountered in first year physics examinations. The typology used recognises four basic task types: "routine operations", "novel problems", "intuitive - interpretive questions" and "bookwork". Thus far it appears that average student performance is usually strongest for routine operations, variable for bookwork, weak for novel problems and weakest for intuitive – interpretive questions. The present phase of the study examines the question of whether the students experience a given assessment task the way the examiner intended – e.g. if a given question was intended by the examiner as a routine operation, do the students experience it as a routine operation or as a novel problem? The answer would appear to depend on the nature and degree of the students' preparation, but the majority of the students responding to our survey were found to agree with the staff – i.e. there was no significant mismatch. Also emerging from this study: a given assessment task need not fit neatly into a single task category; a student can experience a given task as a mix of several operations, each belonging to one of the four types.

1.   Introduction:

Previous studies [1 - 3] have investigated the performance of students in relation to assessment task type and have attempted to find reasons why student performance is weaker in heuristic than in algorithmic problem solving. Additional data now enables us to further develop our question typology and to answer an intriguing question: how reliable is the judgement of an examiner concerning the type of question experienced by the student?

An online literature search failed to provide a suitable ready-made typology – for example the Rhodes' typology [4] was considered but although it was promising, it was not deemed suitable for our purposes. While it is *'extremely well suited for use in Socratic dialogues'* (10) it is intended for content-based questions during the *learning* process with the objective of teaching the students to ask their own questions. Hence, it does not provide a comprehensive set of answers to the question: "Faced with a given *examination* question, what exactly does the student need to *do*, when attempting to generate the demanded answer?" The Rhodes' typology [4] goes some way to giving a list of possible answers to this question but, apart from not being intended for examination questions, it omits at least one important operation: solving the novel problem [5] and hence seems not fully suited to our

purpose. Hence we set about devising our own typology, as follows: A given examination question can require the student to perform a variety of possible basic operations: these are summarised below:

Table 1: Question Typology.

Type:	Abbreviation	Description
"Bookwork"	BW	Material (excluding formulae) has to be recalled from memory and presented, to answer the question.
"Intuitive - Interpretive"	INT	Given information is used to produce a reasoned, possibly non-numerical answer – such as an explanation or a prediction, or a justification etc.
"Routine Operation"	RO	Familiar, well practiced (mathematical) procedures are used to produce the answer.
"Novel Problem"	NP	The student uses a procedure that (s)he has never used before, or else *in a way* (s)he has never used it before, to produce the answer.

This typology is at present very much a work in progress and further developments – such as, the possible inclusion of additional categories, or perhaps the subdivision of the existing categories - will in time be forthcoming.

The question currently under consideration concerns the validity of making statements about student performance in the various question types, such as; "students are good at routine operations" or "students are weak at novel problem solving'. There are two possible complications: Firstly, a student might not experience a given question the way it was intended by the examiner – e.g. an examiner might have intended a question to be a routine operation. A correct answer would normally be interpreted to mean that a student has proven him or herself capable of performing that routine operation. But what if a significant fraction of the students experienced that question as a novel problem instead, simply because they had not prepared sufficiently well to be able to use the usual algorithm? In answering the question correctly, at least some of them would be proving that they were able to solve a novel problem, but not that they were able to perform the routine operation. How would the examiner know the difference? Secondly, a given question might involve more than one (possibly all four) of the categories above. For example, a question which is basically intended as intuitive – interpretive might require the student to recall some piece of material - in which case there is an element of 'bookwork' involved. A student who cannot – through poor preparation – recall a key law, definition or other piece of information, might *for that reason* be unable to produce an answer to what was intended as an intuitive – interpretive question. What can we then conclude about that student's strengths and weaknesses and how do we go about any form of remediation for students in need of help?

An analysis of a first year physics examination at mid-year in 2010 yielded the following figures:

Table 2: Student performance by question type: mid-year 2010.

Question Type: (as deemed by examiner)	Percentage contributed to the total mark	Average student mark
Bookwork	6%	77%
Intuitive - interpretive	4%	18%
Routine Operation	69%	59%
Novel Problem	21%	40%

These figures might possibly tempt one to the following conclusions:
- Students performed the best at bookwork.
- Student performance at novel problem solving clearly was lower than at routine operations.
- The marks were lowest – dramatically so – for the intuitive / interpretive questions.

But are these reasonable conclusions? Certainly, 77% looks different from 18%, even without a statistical test of significance but what about 59% as opposed to 40%? Caution dictates that not only should we perhaps perform such a test, but that first, the two questions mentioned earlier should be addressed: i.e. to what extent would the students agree with the examiners about the classification of the questions? The other being: aren't at least some of the questions of 'mixed' type? Without answers to these questions, what can we conclude from such data?

To attempt to answer these two questions we had a class test written in May this year classified according to question type by: the examiners who set the questions, by other teaching staff who are familiar with the course, and by the students who had written the test. We then set about comparing staff opinion with student opinion.

The course involved was a first year physics course for engineering students. Participants were not randomly selected. This was because:
- The number of suitable staff available was small (n = 4) – we used all available staff with familiarity with the course – i.e. who would have the necessary insight; from having taught on the course - to be able to make a reasonable judgement on question type. The staff participants were thus *purposively* selected.
- The student respondents (n = 177) were students attending the "academic development" tutorial sessions during the week following the writing of the test and were thus effectively self-selected. It should be noted that this group of students represents a range of performance levels from extremely poor to excellent.

Each participant was asked to rate each of the questions in the test as shown in the example (question 2 from test paper) below:

A telephoto camera lens consists of a converging lens with a focal length of 300 mm and a diverging lens with a focal length of – 100 mm, separated by 275 mm. When a photograph is taken, light from a distant object must pass first through the converging lens, then through the diverging lens and then must form a real image on the sensor. How far from the diverging lens must the sensor be situated to form a focussed image?

Question type:	Featured:				
	Not at all	Slightly	Moderately	Strongly	Exclusively
Memory / Bookwork:					
Intuitive / Interpretive:					
Routine operation:					
Novel problem:					

Figure 1: Selection from May test with evaluation grid for survey.

This question was selected for discussion here because it yielded the worst average performance of the test: 8%. From the examiner's perspective this was essentially a routine operation and as this question type has in the past usually yielded good average performances from the students (see table 2), the exceptionally low score actually achieved is puzzling to say the least. Perhaps the students didn't experience this question as a routine operation.

The respondents were asked to consider each question and then to mark the blocks in the grid according to their perception of the question. For example a student who remembered this question as nothing but a routine operation would have marked the block under 'exclusively' in the 'routine operation' row and marked all other types as 'not at all'. If the student felt that there was also an element of, say, interpretation and intuition involved, then the respondent would have marked a non-zero response in this row as well. The responses were given numerical values as follows:

Table 3: Numerical equivalents for responses used in typology survey.

Featuring:	Occurrence Rating
Not at all	0
Slightly	1
Moderately	2
Strongly	3
Exclusively	4

The occurrence ratings provided by the respondents were entered into a spreadsheet for numerical analysis, in which we attempted to assess to what extent the students found themselves doing what the examiners intended while answering the test questions.

2. Findings

Three caveats:
- The respondents were not randomly selected and -
- the numerical values used (see Table 3 above) are ***codes***, not measurements and -
- were based on highly subjective decisions on the part of the respondents.

Hence, we should be cautious about reading too much into the data. Nevertheless, some useful interim conclusions can be drawn.

2.1 Individual questions:
Question 2 (discussed above) yielded the following result:

Table 4: Average occurrence ratings for question 2

Question 2:	Occurrence Rating	
Question Type:	Staff Average	Student Average
Memory / Bookwork:	1.0	1.3
Intuitive / Interpretive:	2.0	1.7
Routine operation:	3.0	1.4
Novel problem:	0.5	1.3

There is a temptation to explain the very low average performance for this question as follows: The students' average rating for 'routine operation' was less than half the value for the staff; also, the rating for 'novel problem' is higher for students than for staff. As past studies [3] have suggested that students are generally better at routine operations than they are at novel problem-solving, this might seem to explain the low average performance of 8%. However a closer look at the numbers suggests that this conclusion might be unwarranted. Firstly with an average performance of only 8% for this question, surely there would have been a *much* higher discrepancy between the staff and student ratings than there is? Also, most of the apparent discrepancies can be made to disappear by as simple an operation as rounding off the average ratings. Perhaps the apparent differences are not so real.

2.2 The test as a whole:

Further analysis involved answering the question: "what percentage of each category of respondents gave each of the various question *types* a given occurrence rating across the test as a whole?" The answer to this question is summarised below:

Table 5: Average overall occurrence ratings for staff and students.

Type	Staff					Students				
Memory / Bookwork:	18%	13%	25%	10%	35%	27%	11%	15%	20%	27%
Intuitive / Interpretive:	38%	18%	25%	20%	0%	41%	12%	17%	21%	9%
Routine operation:	45%	10%	3%	23%	20%	44%	12%	17%	17%	11%
Novel problem:	80%	5%	13%	3%	0%	70%	10%	7%	8%	5%
Occurrence rating:	Not at all	Slightly	Moderately	Strongly	Exclusively	Not at all	Slightly	Moderately	Strongly	Exclusively

The data summarised here seem to show that mostly, the students agree with the staff on what a given test question amounts to by way of question type. Hence, while there are individual differences between staff and students – i.e. there *are* individual students who, for example, find themselves solving as a novel problem a question which the examiner had intended as a routine operation – the overall situation is that the majority of the students agree with the staff consensus about what type of question they are answering.

2.3 Conclusions:

Caution is still needed when drawing conclusions about students' ability to perform according to question type as, apart from the three caveats mentioned above, some examination questions involve the student in more than one category from the typology. In these cases we need to be cautious about what we conclude from students' average performance in a given question. However, it does seem that while there will be exceptions, we can with reasonable confidence regard our own judgement about question types to be reliable. If an examiner deems a question to be a routine operation, the majority of the students will experience it as a routine operation as well.

References:

[1] Clerk D and Naidoo D 2009 *Proceedings of the Seventeenth Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education*, 72-76

[2] Clerk D and Naidoo D 2010 *Proceedings of the eighteenth Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education*

[3] Clerk D and Naidoo D 2010 *55th Annual Conference of the South African Institute of Physics*

[4] Wenning C J, Holbrook T W and Stankevitz J 2006 *Journal of Physics Teacher Education Online,* 4(1) 10 – 13

[5] Martinez M 1998 *Phi Delta Kappan*, 79, 605-609