Termites in our Tests? The role of stigmergy in our examination system

Douglas Clerk, Deena Naidoo and Claudia Albers

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

E-mail: Douglas.Clerk@wits.ac.za

Abstract. This study is inspired by perceived shortcomings in the 'problem-solving' abilities of undergraduate physics students. A detailed analysis of student performance in examinations in relation to the type of question being answered - for a first year physics course for engineering students - has been undertaken. The data collected show that firstly, there is empirical evidence in support of these perceptions. Secondly, evidence has also emerged that there is a favoured question-type that can explicitly be taught and relatively easily mastered, and which typically makes up a sufficiently large fraction of an examination, that students can pass without having to demonstrate any real problem-solving ability. What students need to demonstrate instead is a well-developed ability to expedite routine operations at various levels of complexity – which by definition does not amount to problem-solving. It is possible that this bias has become established *stigmergically* via a feedback process sometimes called 'backwash', to which candidates, examiners and instructors have all been party. Candidates learn what kind of questions to expect, examiners learn what kind of questions candidates can be expected to answer, and instructors learn what kind of question-answering skills need to be taught - by traces left by these agents in the system's environment. The third outcome of the study has been the emergence of a taxonomy of question-types typically set in physics examinations.

1. Introduction and rationale

Problem-solving is an educational outcome of extraordinary importance. Jonassen [1] for example, rates it as "the most important cognitive goal of education". Martinez [2] states that "the most important kinds of human activities involve accomplishing goals without a script". Given this wide recognition of its great importance, one would imagine that problem-solving would surely have been pursued as an educational goal with considerable vigour and hence, success. However this seems not to be the case: Gil-Perez, Dumas-Carré, Caillot, & Martinez-Torregrosa [3] regard the "abundance of literature" on problem-solving as evidence not only for "the relevance of problem-solving to the learning of science", but they also see it as "evidence of a general failure of students at problem-solving tasks". The rationale being, that if there were no 'problem' with problem-solving, there would not have been such a great abundance of literature concerning it. Apart then from 'tea-time talk' amongst disgruntled lecturers, there is thus considerable literature support for the view that problem-solving lacks general success as a learning outcome, for example: "On routine problems – that is, problems that are like those they have already learned to solve – they excel; on non-routine problems – i.e., problems that are not like any that they have solved in the past – they fail. Similar examples can be found in other academic domains, including reading and writing" [4].

This lack of success of problem-solving as an educational outcome is somewhat curious and indeed, of some concern. Hence, we have set out to shed some light on the situation and attempt to suggest a mechanism for what is at play. This study examines the role of the examination system, in the development – or perhaps the *failure* to develop – of problem-solving competencies in students of physics. Three questions come readily to mind, the first being *what* is 'problem solving'? The other two serve as our research questions: is there empirical evidence of a deficiency in problem-solving abilities among our students - and if so, what could be causing the deficiency? These questions have been addressed by an on-going analysis of examination and test questions in physics, and of the marks (grades) students typically achieve when answering them, spanning now several years.

The first of the three questions above concerns our working definitions – without which further debate could prove to be futile. Apart from a dictionary, the obvious place to look for such definitions is of course the literature. This has proved to be something of a 'problem' as explicit definitions are not always given in the literature. This fact has itself been commented on – for example: "researchers into paper and pencil problem solving do not generally raise the question of what constitutes a problem" [3]. Where the question is raised, there is sometimes significant – but usually *not* exact – agreement with our own understanding. We have been able then to select a sub-section of the relevant literature (see below) in which the given definitions are not only explicit, but are also both clear and succinct, *and* which resonate well with our own understandings; and we develop our own working definition based on those, as follows:

Tasks in physics – both for training and assessment - can often be performed by means of an algorithmic process. An algorithm - here defined as a series of instructions to be followed sequentially in the performance of some task (i.e. a procedural 'script') - can be learnt, and could thus become a routine task such as those referred to above by Mayer [4]. Note here that Mayer does not make the same semantic distinctions as we do: he refers to both routine and non-routine tasks as 'problems'. Nevertheless the claim he makes supports our findings and, in a later publication, Mayer [5] does show that he is at least *aware* of 'our' meaning.

Any student then expediting such a learnt algorithm in response to an examination question might indeed be 'performing well' - in the sense that good grades would be awarded for their efforts - but would not be problem-solving as, by our chosen definition, a task is not a 'problem' if its algorithm is already known. This follows the views of authors such as Adams & Weiman [6], Bodner [7], and Martinez [2]. Problem-solving occurs not when the student generates the answer to a question by expediting a familiar, pre-fabricated algorithm, but rather, when the student *constructs* the necessary algorithm. Hence, our central working definition becomes: ***A problem is a task with a cryptic algorithm*** – i.e. an algorithm not initially known to the solver but that *becomes* known during the course of and through the process of solving the problem. The point of 'problem-solving' is then to discover the algorithm.

Evidence exists – both in the literature, [3, 4] and in empirical form - from this study at least - that the problem-solving deficiency referred to here is indeed real. We also suggest a possible causal mechanism, known as 'backwash' which is documented in the literature [8, 9]. It consists of a feedback, or perhaps more accurately, a feed-*forward* loop in which past assessment influences current and future teaching, learning, and assessment. Backwash has been quite extensively researched in the learning of languages but less so in physics, although the continuing demand for past examination papers suggests that it is very much alive and well as a contributing factor.

2. Theoretical Framework

A theoretical framework provides an ontological and epistemological context within which to design and conduct research, and hence to interpret findings [10]. In educational research generally, such a framework must of necessity be somewhat complex and must draw on several sources, as it involves complex issues where effects usually arise from multiple causes, and where we deal with, as Redish [11] says, "a strongly interacting many-body system in which observations change the system in uncontrollable ways" – i.e. a *complex* system.

What is of concern in this study is essentially a pattern of learnt behaviour, both individually and collectively – hence, constructivism as a theory of learning would broadly underpin the research. We are interested in the collective behaviour of individuals in society when learning to cope with a difficult situation – i.e. that occasioned by 'high-stakes' examinations – which we see as any examination with significant consequences [12] for the examination candidate. A candidate's need to perform well varies in intensity depending on how high the 'stakes' are. Indeed failure can have life-altering consequences for the candidate, and hence would also be of some importance to the instructor and also the examiner. There is thus high pressure on all agents acting within the system towards the achievement of 'successful' outcomes. We believe that what we describe here is a negative and unintended consequence of this pressure [12, 13].

We draw thus on the theoretical perspectives of constructivism, metacognition and the basics of complex systems theory, to suggest that students and instructors, being agents acting within a complex system, will construct a '*meta-knowledge'* that allows them to make sense of and act successfully within that system – in short: knowledge of how to pass an examination and what it *means* to do so. This strongly suggests a link with metacognition, as the meta-knowledge is not *subject* knowledge. Rather, it is knowledge about how to learn the discipline in such a way that one can satisfy the requirements of the examination system – i.e. that which a student would need to know in order to be a successful strategic learner [15].

The active construction of knowledge in and by the mind of the learner involves the production of viable explanations for our experiences for better navigation through what we commonly call 'reality'. Ideally, the mental models that we thereby create [16] would as far as possible need to be a true reflection of that reality, in order to function as reliable navigational tools. This knowledge is not constructed in isolation hence it is necessary also to recognise the strong input of the social environment in the construction of this knowledge. The meta-knowledge here considered would be constructed as a *collective* 'effort' within a community – that comprising principally the students, instructors and examiners of physics; and also, but more peripherally, the various other stakeholders.

Collective efforts within any system require some mechanism of coordination between the various agents and we are informed therefore also by *stigmergy*, as the agents in the system both make and sense 'signs' in the environment of that system, that influence the behaviour of other agents [17, 18, 19]. These signs, we believe, consist on the one hand of the discernible patterns of question-types found in the examination papers and on the other, of the performances – reflected in their marks (grades) - of the examination candidates who answer them.

3. Methodology

Our basic aim was to assess the extent to which examination questions typically require students to work in 'algorithmically unfamiliar territory' and to compare their performance in such questions with that in other types of questions, in particular ones involving familiar algorithms. This we did by identifying essentially what it is that the student would need to do in order to produce the answer to a given question, and then to assess the average performance of the students for each question. This approach has allowed us to identify several question-types typically asked in (our) physics examinations and hence, to assess how the students perform when answering each of the various question-types.

In order to link student performance with question-type, we recorded the marks awarded to each student for each individual question on a spread-sheet (see Figure 1 below). This allowed the convenient calculation of an average mark for each question, to indicate the group performance. Data was thus captured from examinations given to a first year course in 'general' physics for engineering students. We base our claims here on the data from both the mid-year and the final examination results of 2012 and 2013. We present below a sample spreadsheet for a class test, showing a selection of the data to illustrate the capture and analysis of the data.

Question:	1		2		3		4				
Subsection:	a	b	a	b	a	b	a	b			
Sub-subsection								i	ii	iii	Total:
Out of:	10	8	1	7	8	4	3	3	4	2	50
Student #	4	8	1	7	7	2	3	3	4	2	41
Student #	2	1	1	0	2	0	1.5	3	0	2	12.5
Student #	4	0	0	2	1	1	1.5	0	2	0	11.5
Student #	5	8	1	2	5	0	3	3	0	0	27
Student #	10	5	1	4	3	0	3	3	4	2	35
Student #	10	8	1	2	1	0	3	3	2	0	30
Student #	10	5	1	3	1	1	3	3	4	0	31
Student #	4	8	1	4	7	1	3	3	4	2	37
count:	832	832	832	832	832	832	832	832	832	832	
Sum:	4957	3829	665	3111	3146	845	2399	2222	2524	728	
Average:	6	5	1	4	4	1	3	3	3	1	
Average %:	60	58	80	53	47	25	96	89	76	44	
Recall & Present:	1	1	4	1	2	1	4	1	1	1	
Intuitive/Interpretive:	2	1		3	1	3			2	3	
Familiarity:	3	3		1	3	1		4	3	2	
Complexity:	2	2		3	2	2		0	2	2	
Topic:	Optics	Optics	Fluids	Fuids	Waves	Waves	Thermal	Thermal	Thermal	Thermal	

Figure 1: Sample spreadsheet illustrating types of data emerging from study.

Limitations of space allow the marks of only eight of the original class of more than eight hundred students to be shown, but the averages shown are calculated from the full data set.
A five point typology coding scale – see Figure 1 - was used, ranging from zero to four.
The following points illustrate their use in this study:

- Two questions (question 2a and question 4a) were rated '4' for recall – this means that the questions were ***pure*** 'recall-and-present' questions. The blanks in the blocks below indicate that there was no computation involved, hence the familiarity or complexity of an algorithm was simply 'not applicable', as was the use of 'physics intuition'. Both questions yielded very high averages – eighty percent and ninety-six percent respectively.
- Three questions (2a, 3b and 4biii) were coded '3' for 'physics intuition'. Two of the three yielded failing averages, one of which was the lowest average – twenty five per cent - in the data set. This question-type has stood out both for its relative rarity and the extremely poor performances usually associated with it.
- Question 4b (i) was a computation and was rated a '4' for algorithmic familiarity and a '0' for complexity, and yielded the second highest average of eighty nine per cent. All other 'algorithmic' questions were given lower familiarity ratings and also yielded poorer performances.

4. Outcomes & Results

A taxonomy of question types was thus established by considering what the student would need to ***do*** to produce an answer to a given question - as follows:
a) Recall: either ***recall-and-present*** – such as laws, definitions and sundry 'bookwork'; or else ***recall-and-use*** where material is recalled and then utilised during the production of other parts of the answer - see 'computations' below.
b) Intuitive/interpretive (questions): using a previously constructed mental model - largely *qualitatively* - to make a judgement or prediction, or else to explain something - sometimes called 'physics intuition'.
c) Computations: – questions involving mathematical operations, being either 'routine operations' or 'novel problems' - which mark the two ends of the same continuum; and constituting one dimension of the "difficulty" of the problem-solving process. In a routine operation the algorithm is recalled and used, in a novel problem the algorithm is discovered and constructed during the solving process.

Further we note that:

- As we have reported before [20], several of these taxonomic types often feature as ***elements*** in a question, rather than being the sole attribute of a single question.
- The 'routine operation' emerges as the clear favourite question-type by a considerable margin (see Figure 2 below) and,
- Students perform reliably well when dealing with familiar algorithms – i.e. in routine operations - even where the algorithm is quite complex (see Figure 3 below).

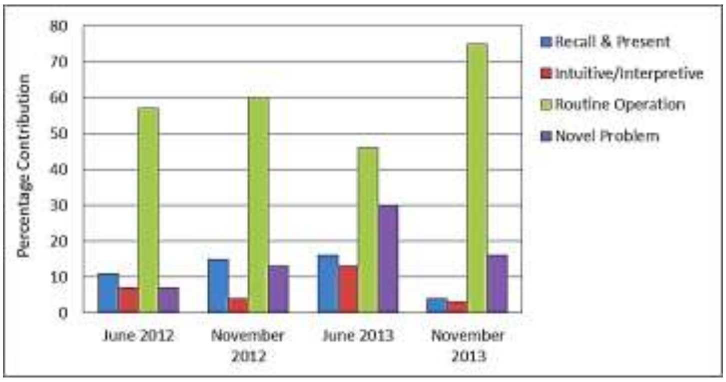

Figure 2: Percentage contribution to examination marks by question type.

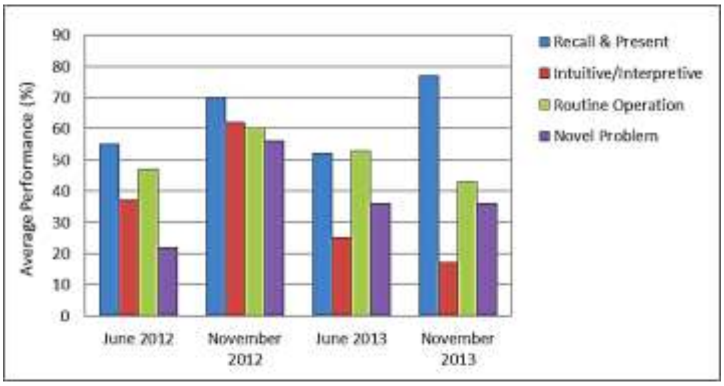

Figure 3: Average performance – as percentage - by question type.

5. Concluding discussion

It was thus possible for students to pass any of these examinations in the almost complete absence of true problem-solving. To support this claim we point out that in the papers analysed, 60% to 70% of the marks were allocated to questions demanding only a working knowledge of some routine operation. The majority of the students who passed would have scored between 50% and 70% - i.e. within the same range. Students who achieve anything up to 70% thus provide no evidence that they are capable of much other than performing routine operations. Given the widely agreed importance of problem-solving as a learning outcome [1], there is surely something amiss with this situation.

Backwash feeds the message into the system that routine operations are the dominant question-type and that a candidate can expect to pass an examination knowing almost nothing else. The medium of this 'message' is essentially stigmergic – consisting of the patterns discernible in the examination questions and the performances of the candidates answering those questions. These patterns are analogous to the pheromone trails left in the environment by foraging ants or nest-building termites.

Motivation for teachers to maximise their pass rates above and beyond the ‘satisfaction of a job well done’ is provided by the ‘performativity’ ideology of educational management now prevalent in most countries [21, 22] in which target metrics (e.g. pass and distinction rates) that should serve as management tools, indicating the achievement of more fundamental learning outcomes [23] become instead, goals in their own right that instructors and examiners must strive to achieve, or face sanction.

Given the definitional vagueness concerning what constitutes problem-solving found in the literature, [3] instructors could well be teaching routine operations under the impression that this amounts to the teaching of problem-solving. As good pass rates can be achieved by teaching routine operations and not much else, we suggest that this may be why the ability to solve true – i.e. novel – problems is so underdeveloped. If we exclude true problem-solving from our examination system, we in effect exclude it from our *de facto* curriculum. This omission may be a significant, but under-recognised ingredient in the education ‘crisis’ from which our country is currently suffering.

A parallel analysis of the South African school-leaving (physics) examination since 2008 reveals that here also, routine operations are the dominant question-type [24]. Unfortunately data for student performance *per question* are not available for these examinations.

References

[1] D. H. Jonassen, (2010) *The 11th International conference on Education Research New education paradigm for learning and instruction*, , pp. 1–15.

[2] M. E. Martinez, *Phi Delta Kappan*, vol. **79**, no. 8, pp. 605–609, 1998.

[3] D. Gil-Perez, a. Dumas-Carré, M. Caillot, and J. Martinez-Torregrosa, *Stud. Sci. Educ.*, vol. 18, no. 1, pp. 137–151, 1990.

[4] R. E. Mayer, *Instr. Sci.*, vol. 26, pp. 49–63, 1998.

[5] R. E. Mayer, *International Encyclopedia of Education*, 2010, pp. 273–278.

[6] W. K. Adams and C. E. Wieman, *AIP Conf. Proc.*, vol. 883, pp. 18–21, 2007.

[7] G. M. Bodner, *Univ. Chem. Educ.*, 2003.

[8] L. Cheng, *Language and Education*, vol. 11. pp. 38–54, 1997.

[9] Q. Xie and S. Andrews,*Language Testing*. 2012.

[10] L. Mack, *Polyglossia*, vol. 19, pp. 5–11, 2010.

[11] E. F. Redish, *Proc. Int. Sch. Phys. Enrico Fermi Course CLVI*, p. 63, 2004.

[12] N. Noddings, *Theory Res. Educ.*, vol. 2, no. 3, pp. 263–269, 2004.

[13] G. J. Cizek, *Educ. Meas. Issues Pract.*, vol. 20, pp. 19–27, 2001.

[14] T. G. Lewis, *Cogn. Syst. Res.*, vol. 21, pp. 7–21, 2013.

[15] J. B. Biggs, *Teaching for quality learning at university: what the student does.*, 2003, pp. 11–33.

[16] D. Clerk and M. Rutherford, *Int. J. Sci. Educ.*, vol. 22, no. 7, pp. 703–717, 2000.

[17] F. Heylighen, *Cognition*, pp. 1–23, 2011.

[18] F. Heylighen, in *Lewis, Ted & Marsh, Leslie (eds), Human Stigmergy: Theoretical Developments and New Applications (Studies in Applied Philosophy, Epistemology and Rational Ethics*, 2015, pp. 1–43.

[19] H. Van Dyke Parunak, *Proc. 2nd Int’l Conf. Environ. Multi-Agent Syst. II (E4MAS '05)*, vol. 3830, no. 2005, pp. 163–186, 2006.

[20] D. Clerk, D. Naidoo, and S. Shrivastava, *Book of Abstracts, 57th Annual Conference of the South African Institute of Physics.*, 2011, p. 85.

[21] S. J. Ball, *J. Educ. Policy*, vol. 18, no. 2, pp. 215–228, Apr. 2003.

[22] S. J. Ball, *Br. J. Educ. Stud.*, vol. 60, no. March, pp. 17 – 28, 2012.

[23] J. Hennessy and P. M. McNamara, *English Teaching: Practice and Critique*, vol. 12, no. 1, pp. 6–22, 2013.

[24] D. Clerk and D. Naidoo, *Proceedings of the 18th Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education*, 2010, pp. 142 – 147.