Death of an outcome – the role of stigmergy in our examination system

D P P Clerk [1,2] **and D Naidoo**[1]

[1]School of Physics, University of the Witwatersrand, Johannesburg, South Africa, Private Bag 3, WITS, 2050
[2]douglas.clerk@wits.ac.za

Abstract: This paper reports on an analysis of questions in physics examination papers at matriculation and first-year university level. The intention is to throw light on an on-going concern, sometimes expressed by university level physics instructors, that there is a general weakness in problem-solving ability among their students. This begs several questions; the first and most obvious one being what we mean by 'problem-solving'. The second is whether this perceived weakness has an objective reality. Evidence supporting this can be found in both the literature and in examination results. The third question asks what might be causing this state of affairs. To attempt to answer this, examination papers at both matriculation and university level (first-year) were analysed with a view to finding what *types* of questions are typically asked, and to ascertain whether there may perhaps be a bias in favour of - or against - specific types of question. The evidence shows that there is indeed a favoured question-type that can be explicitly taught, and relatively easily mastered - and which typically makes up a sufficiently large fraction of an examination that a candidate can pass the examination without having to demonstrate any real problem-solving ability. What an examination candidate *does* need to demonstrate is a well-developed ability to expedite routine operations at various levels of complexity, which - depending on one's espoused definition - does not amount to problem-solving. The bias away from novel (or heuristic) problem-solving is here seen as a stigmergic sign that indicates to candidates, instructors and examiners alike that the ability to deal with novelty, which is the essence of problem-solving, is not needed in order to pass - or even to do well in - these high-stakes examinations. Hence the need either to teach or to learn problem-solving could come to be seen as unnecessary. This state of affairs is repeated - and hence reinforced - at each successive annual iteration.

1. Introduction

1.1Rationale

Problem-solving is widely held to be of paramount importance as an educational outcome [1], especially in physics. Given this, concerns that are sometimes raised about the poorly developed 'problem-solving' skills of new first-year undergraduate physics students – some anecdotally and some in the literature [2] - beg three questions: a) what exactly is '*problem solving'?* b) is there empirical evidence supporting the reality of students' supposedly deficient problem solving skills? and c) if so, what could be causing the deficiency? These three questions have been addressed by an analysis of examination questions spanning several years, to see if features inherent in the examination system might perhaps be contributing to the deficiency. The examinations targeted for scrutiny were the South African matriculation examination in physics, and the first-year examination for a physics service course catering for engineering students at a South African university.

The investigation has produced evidence suggesting not only that the problem-solving deficiency is indeed real but also a possible causal mechanism. The candidate mechanism is actually well known and documented, and has been called 'backwash' - among other names - in the relevant literature [3], [4] and consists of a feedback loop in which past assessment influences current and future teaching and learning, and hence also future assessment. It can further be argued that backwash is an instance of stigmergic coordination [5] as the information that is fed back into the system is not intended by the examiners to inform - rather it is the very structure of the examination papers that constitutes the information.

In the current study, each sub-section of each question in a given examination was classified according to a typology concurrently under development. The most frequently used question-type has consistently been the 'routine operation' - by contrast, questions demanding true problem-solving have regularly accounted for a significantly smaller fraction of each examination. The consistent pre-eminence of a specific question-type is believed to act as a stigmergic trace in the 'environment' that feeds the message back to students and instructors alike that the most frequently used - and hence, presumably most important - type of question is the 'routine operation', as it is here that the bulk of the marks are typically awarded. Hence, candidates can perform well or at least pass comfortably by focusing on this question-type, possibly to the exclusion of all others. Limited - if any - proficiency at true problem-solving is necessary for success in these examinations, hence there is no real necessity either to learn it, or to teach it – could this be why problem-solving is an underdeveloped skill?

1.2 Working definitions

The question of what is meant by terms such as ***problem*** and ***problem-solving*** must be addressed in order to establish working definitions of these (and other) key terms to enable meaningful debate. The working definitions espoused in this study have been adopted - and adapted - from the literature [3], [5], [6] and [7]. Beginning with problem; the various meanings can be grouped into two broad categories. In the first, a 'problem' is something wrong or objectionable in a state of affairs to which we may be referring. This is obviously irrelevant to our purpose. In the second category, the problem is something perplexing, puzzling or difficult: when we refer to some task as 'a problem' or as 'problematic', we mean that there is some aspect of the task that makes it to some significant degree difficult to perform the task. This suggests that problem-solving consists essentially of removing, circumventing or otherwise overcoming the 'difficulty'.

Many tasks in or associated with the teaching and learning of physics can be performed by means of an algorithmic process. We argue that such tasks only become 'problematic' when the algorithm is not known *a priori* to the solver. This follows the views of authors such as Bodner, [5], Martinez [8] and also Schuster [10] who all argue that problem-solving does not occur when the candidate produces the answer to a question by expediting a familiar, *pre-fabricated* algorithm - regardless of its complexity. Rather, it consists of choosing, discovering, modifying, devising or otherwise ***obtaining*** the algorithm, prior to expediting it. Problem-solving is 'what you do when you don't [yet] know what to do' [9] (p 2) – i.e. it consists of ***finding out*** what to do. There appear to be no competing definitions in the literature – at least none that address the contrast between what is and what is not a problem. Definitions that *are* provided often mean essentially the same as that espoused in this study. However quite commonly, no definition is offered at all; perhaps on the assumption that none is needed – that '*everybody knows*' what is meant. This fact is sometimes commented on in the literature: according to Gil-Perez et al, [12], themselves citing Garrett, [13] "researchers into paper and pencil problem solving do not generally raise the question of what constitutes a problem" (p 138). Garrett himself speaks of a narrow interpretation of the concept (of problem-solving) "by teachers and investigators alike as simply the carrying out of sterile exercises" [13] (p. 70) which suggests that a problem is quite widely seen as simply *any* task or exercise that must be performed, regardless of what its performance might actually involve. This view is unsatisfactory: unless problem-solving is somehow distinct from the answering of other types of question, and that distinction is made clear, there is little point in further debating or investigating the matter.

Thus, while there might be general agreement that to be a 'problem', a task needs to be in some way 'difficult' for the solver – what is argued here is that what needs to be difficult about the task - that which *makes* the task 'problematic' - is that whatever algorithm may be (or may *become*) associated with performing the task is at the outset at least partially unknown to the solver. The critical element then is the initial absence of a fully-known algorithm rather than such features as the length and/or number of steps in the process, or else the style of wording in the task question. This provides a definition in which there is a clear and qualitative difference between tasks which are problems and those which are not - a difference in *kind*. Were the distinction to be based instead on the length and or complexity of the algorithm concerned, the difference would be one of *degree* – which would raise the inevitable and difficult question of precisely *where* to draw the line between what is and what is not, a problem. Hence, the working definition of 'problem', appropriate to typical tasks in physics, that we have adopted is: ***A problem is a task with a cryptic algorithm***.

As it is so central to our definition of 'problem', some discussion of our meaning of 'algorithm' is now appropriate. In its broadest sense an algorithm is ***a series of instructions to be followed sequentially in the performance of some task*** [12]. An algorithm can range from trivially simple, to sufficiently extensive and complex that a computer would be needed to expedite the 'solution'. Thus even the mathematical instructions embodied in the formula for some quantity effectively constitute an algorithm.

2. Methodology

Having adopted these definitions, the next step in the process was to analyse examination papers in an attempt to establish the extent to which the questions in them required candidates to be working in 'algorithmically unfamiliar territory'. This broad approach has also allowed another question to be addressed – i.e. what *are* the various types of questions typically asked in (our) physics examinations, and thence, how many marks are typically awarded to each type?

2.1 Question-type taxonomy

The taxonomy of question types was essentially established by considering what the examination candidate – in particular the '*prepared* candidate' - would need to do in order to produce an answer to a given question. The prepared candidate is one who possesses the necessary mental capacity – including both inclination and aptitude - to succeed in the academic endeavour being undertaken. The prepared candidate is one who has 'engaged' with all the necessary coursework – i.e. has made diligent and mostly successful attempts to master all of the necessary skills and to construct all the necessary knowledge. Most importantly, this candidate will have made diligent efforts to prepare adequately for the specific examination in question. (In other words, this is the candidate who 'should' be able to pass the examination.) Any candidate not fitting these criteria can arguably be expected to have at least *some* difficulty answering questions in a given examination. It should thus not be in any way surprising if an '***un***prepared' candidate were to have difficulty answering a given examination question, and while this eventuality is of obvious significance for gate-keeping purposes, it may not necessarily signify anything about the merits – and especially about the *nature* - of a given assessment task. In fact, it is more likely an indication that the assessment task has indeed 'done its job' in identifying a shortcoming in the candidate. There would surely be something wrong if a truly unprepared candidate *were* able to answer enough questions to pass an examination. Conversely, if a *prepared* candidate experiences difficulty with a given assessment task, this *may* indicate something significant – perhaps that the task itself is unreasonably difficult, or else that the task is testing some skill that this specific candidate has not mastered – which would suggest that the candidate might actually *not* have been so well prepared after all.

Hence, in the classification of a given question – or an element of a question - the central consideration was: *what would the prepared candidate have needed to* ***do*** *in order to produce the answer to this question*? Answering this question must be based on the examiner's intention - or else on an experienced instructor's judgment. Most other possible criteria, such as the candidate's state of mind during the examination, are simply beyond our reach, even though they may arguably be far more valid. Interviewing candidates may be the best way to elicit their thought processes but has several serious limitations. For example, one obviously can't interview the candidates while they are busy writing, even though this would be the best time from a data-gathering perspective. Interviewing them subsequent to the examination gives them time to forget how they were thinking *at* the time of writing and also for previously novel procedures to have since become familiar. The number of candidates one *can* interview – at any time - is also extremely small for logistical among other reasons

From this, the following question 'typology' has emerged:

Recall: Recall can feature in two basic ways; either the candidate is required simply to show that s/he *can* recall some item deemed by the examiner to be important, or else the candidate needs to recall something *and then apply it* when answering other parts of the question. These two types are called ***recall-and-present*** and ***recall-and-use*** respectively. Recall-and-present questions can range from the trivial; such as "state Newton's third law" - to the quite complex, for example being required to present a memorised derivation, such as that for Snell's law. In the recall-and-use question, there can be quite a wide variation – including but not limited to sign conventions - in what the recalled item is used for.

Intuitive/interpretive (or perhaps '***insight***') questions: These are essentially non-algorithmic questions as the candidate is not expected to use any overtly algorithmic procedure – i.e. there is no requirement to perform a calculation; indeed, numbers are often not given. What is involved instead is what we might call ***analogue*** reasoning (i.e. as opposed perhaps to ***digital*** reasoning). The candidate needs to 'run' a previously constructed mental model of the situation featured in the question in order to produce the answer to the question posed. The candidate could be required to make some judgement or prediction - or else to explain something described in the question. To do this the candidate might also use ***proportional*** reasoning. To illustrate, the candidate might be asked to explain why a bat might perhaps misjudge distances due to variations in air temperature – akin to our experience of altered depth perception due to refraction when using a diving mask underwater.

Computations – i.e. questions demanding mathematical operations of one kind or another - are either 'routine operations' or 'novel problems' - or something in between. Any calculation where the candidate knows – or at least *should* know – the algorithm well, from repeated previous exposure is a routine operation. Algorithms become known to the candidate through drill and practice and thus any algorithm regardless of its length or complexity is at least *potentially* a routine operation. By contrast, the **Novel Problem** is a task where at least some portion of the algorithm – if not the whole of it - is not already known to the candidate, and must thus be devised by the candidate as part of the exercise.

Rather than being two utterly different processes, routine operations and novel problems can be better seen as the two ends of the same continuum. A problem-solver's familiarity with the relevant algorithm determines where on the continuum he or she is operating when solving a given 'problem'. The solver might have complete prior knowledge of the necessary algorithm and thus be operating at the one extreme, or else the solver could start off completely ignorant of the algorithm and would therefore be operating at the other. Usually, the solver would be operating somewhere in between the two extremes, with partial prior knowledge of the necessary algorithm. There would of necessity be some degree of *creating* a usable algorithm – perhaps by modifying one or more existing algorithms. For convenience, we could refer to this dimension as the "familiarity" – or else conversely, as the "novelty" - of the algorithm.

Given enough exposure and training, any novel problem can and eventually will become a routine operation - for a given candidate, the novelty of a task is a matter of his or her individual experience and this will change with time.

2.2 Coding

Having established the various types of question, the next step was to assign coding numbers to the questions, according to a scheme that is best illustrated with an example. To begin with, each subsection of each question was assessed with respect to its question typology and the estimates of the contribution made to the allocated marks by each question-type element were tabulated. The rationale for this process is illustrated below.

Table 1: Illustration of the 'Anatomy' of a typical coding table.

NSC Examination November 2008		**Question: 5.1**	**5.2**	**5.3**	**Mark Contribution**
Question-Type Element	**R&P**		1		2÷4 = **0.5**
	R&U	1		2	(4÷5)+(3÷2) = **2.3**
	I&I		3	2	(2÷4×3)+(3÷2) = **3.0**
	RO	4			4÷5×4 = **3.2**
	NP	0			**0**
Marks Allocated		**4**	**2**	**3**	**9**

Sub-question 5.1 Asked candidates to calculate the speed of an initially stationary car immediately after another moving car has collided with it. The initial speed of the moving car and both masses were given. This question was judged to be principally a routine operation, with some degree of recall needed outside of the algorithm *per sé*. Hence it was rated as 1 for R&U (recall and use) and 4 for RO (routine operation). The other sub-question columns were populated in like manner. For the column headed 'mark contribution' the calculation shown rested on the following reasoning: there was one instance where the candidate was required to recall and present material as a minor part of the production of the answer – the rest being a process of reasoning, hence the ratings for this sub-question as 1 for R&P and 3 for I&I. The sub-question was allocated 2 marks overall and hence recalling and presenting represented 25% of what the candidate was deemed to have been doing. The overall 'contribution' for R&P was thus calculated as $2 \div 4 = 0.5$. In the next row down, recalling and using was given a '1' for sub-question 5.1 and a '2' for sub-question 5.3 - hence the calculation $(4 \div 5) + (3 \div 2) = 2.3$. The rest of the table (and also each of all the other tables) was populated in like manner.

2.3 Inter-Coder Reliability

The question of subjectivity and inter-coder reliability must be addressed with some seriousness. Given that the coding was largely a matter of exercising judgement – largely based on the coder's insight and experience - how can there be any confidence that other coders might not have perceived any given examination question differently? It is argued on the basis of the claims that follow, that there is in fact good reason to trust the results here presented.

Claim 1: The coding process was actually quite mechanical, amounting largely to fitting the process of production of an answer to a question-type definition. For a first example, consider a question that required a candidate simply to recall some memorised item, such as a definition or a law, and then to write it down in response to the question. There is little possibility that this would be deemed to be any kind of a 'problem' other than by a bizarre misinterpretation of - or disagreement with - the espoused definitions of the question types. This is plainly a 'recall-and-present' question and as it asks nothing more of the candidate, it would be awarded a four on the rating scale to indicate that there was nothing else involved. A second illustration involves the case of a question (at matriculation level) which required the candidate to perform a calculation using an equation given in the 'data sheet'. Here again, there is little possibility of significant variation between coders. To answer the question, the candidate must use a given equation to perform a calculation that would not differ in essence from those that have been asked in the past – i.e. a routine operation. The candidate is not required in any sense to construct the required algorithm, hence there is effectively no novelty involved. In the case of the first-year examinations there was no actual 'formula sheet' – students were expected either to derive or – more commonly - to memorise and recall their equations. However, equations *were* quite frequently given, not in a 'formula sheet' as such, but in individual questions, in cases where an equation was deemed by the examiner to be excessively complicated, and hence difficult to recall correctly. This has been standard practise in this course for many years. It also means that candidates are expected to perform most of their calculations using equations that were either recalled from memory, or else were given.

Claim 2: Informal consultations were held both with other examiners and with the chief examiner of the university-level course with respect to their 'intentions' for the various examination questions they set. The issue of what would constitute each of the question types was also discussed frequently with both the supervisor and co-supervisor of the study.

Claim 3: A ‘mini-study’ (14) was conducted, which assessed the perceptions of both students and instructors of the nature of assessment tasks. The intention behind this study was to explore the possible differences between the perceptions of various coders – some being students and others being teaching staff of the relevant university course. Contrary to expectations, the study actually found an unexpected lack of significant differences between the perceptions of students of the relevant course and of physics teaching staff, arising from the data collected. The majority of the respondents perceived questions in quite similar ways, which suggests that great variation between coders is actually not to be expected. This study, which looked at the question of *perceptions* of assessment tasks by the various participants, also addressed the credibility of the question typology.

Claim 4: Personal experience and resulting skill development – the corresponding author has been involved in various roles in teaching and assessment since first starting (high-school) teaching in the late 1970s and hence has considerable experience to draw on – having for example, *set* internal matric-level examinations – as a matric teacher on one of the ‘*project* schools’. This author has in addition, marked matric examinations – both public and internal - on numerous occasions and has also been involved in the setting and marking of first-year level examinations – most significantly as part of the examining panel and the teaching team for the course that was the source of the first-year examinations for this study. This extensive experience – which now spans over four decades - suggests a fair degree of reliability in coding.

3. Outcomes & Results

Firstly, questions cannot always be classified as being purely of a single type. What is far more usual is that there will be several ***elements*** of the taxonomic types identifiable in a given question – in one question the candidate could be required to recall something, to run a mental model, and to expedite an algorithm. The algorithm will be of a given level of complexity and may either be extracted ready-made from the candidate’s repertoire, or else be devised on the spot. Hence coding numbers as discussed above for several categories could potentially be given to a single question.

Secondly, as mentioned earlier, there was a definite bias in favour of one particular question type. Presented in the table below is a distillation of the data extracted from the analysis of the examinations papers, showing the average ‘percentage contribution’ of the various question types spanning the full duration of the study – i.e. from 2008 to 2014 in the case of the NSC examinations and including both the mid-year and final papers for the first year examinations.

Table 2: Overall percentage contribution averages for the various question types arising from the study.

Question Type Element	Average mark contribution as % of the total	
	NSC	First Year
R&P	14.3	9.6
R&U	11.7	15.1
I&I	23.1	15.6
RO	48.4	53.5
NP	2.2	6.2

What stands out is that the ‘routine operation’ (RO) is clearly the favourite question-type, the other types trail behind by a substantial margin, with the ‘novel problem’ (NP) constituting by far the smallest fraction of the mark allocation.

4. Concluding discussion

With the approach of any test or examination, candidates commonly enquire of their instructors about the whereabouts of the ‘past papers’. In South Africa, and presumably in other countries as well, there are thriving businesses whose stock-in-trade is past papers: in South Africa, the name ‘Physichem’ comes to mind quite readily. Past papers are consulted by candidates who are anxious to find out what to expect for high-stakes examinations such as their final school-leaving examination. These investigations inform examination candidates about how to concentrate their efforts to maximum effect during preparation. The candidates are not the only parties so engaged: instructors concerned about their candidates’ performance (which will reflect on themselves) might also scrutinise past papers for the same information. New instructors, faced with setting examination questions of their own may be directed to look at past papers to gain an idea of what ‘sort’ of questions to set. If the extant question papers exhibit a strong bias toward routine operations at various levels of complexity, then why should that not be what candidates will predominantly prepare? Likewise, instructors are thus influenced predominantly to teach and examiners predominantly to set routine operations. If the stigmergic traces – i.e. the marked preference for routine operations and the almost complete absence of novel problems - in the past papers direct candidates, instructors and examiners alike consistently away from ‘true’ problem-solving, why then should any attention be paid to developing the ability to deal with novelty during the teaching and learning process? Ironically, this is somehow not even ‘wrong’ – arguably any examination *should* be set to find out what the candidates have learnt from the course in question. If a course has required a candidate essentially to learn a repertoire of routine operations, then perhaps that is *exactly* what ‘should’ be examined – otherwise there would be a lack of alignment [3] between what is taught and what is examined.

However, most physics courses overtly list ‘problem-solving’ as highly important – if not foremost - among their envisaged learning outcomes. This being so, there would definitely be something wrong if it were possible – and we believe we have demonstrated that it *is* possible, perhaps common – for candidates to pass a physics examination in the complete absence of true problem-solving. If the bulk of a high stakes physics examination demands essentially a working knowledge of routine operations, albeit at various levels of complexity, candidates can then pass ‘comfortably’ without having provided any evidence that they can do much else. But what about those few who can perform above the level of a ‘comfortable pass’? Given that any algorithm can and eventually will become familiar to the diligent and well-prepared candidate through repeated exposure and persistent effort, to what extent does even a high-scoring candidate actually demonstrate problem-solving ability? The exceptionally capable and well-prepared candidate quite probably finds little or nothing unfamiliar in the ‘usual’ type of physics examination – every conceivable operation will, one way or another, have been learnt and then converted into a familiar algorithm by the time that candidate sits for the examination. (Some might argue that this is the whole *point* of ‘studying’!) There is however, simply no way of knowing whether any *true* problem-solving occurred during the examination or not – indeed it probably didn’t – but should we not be concerned about this? Are we comfortable with knowing that even our distinction candidates may have done nothing but demonstrate a superior ability to expedite known algorithms? They have not proved that they can solve a novel (true) problem and well may be entirely incapable of solving one - and if we have taught and examined predominantly routine operations, why should we be surprised?

A mystery is why any well-intentioned instructor would buy into what has been described above. Feedback loops – especially of they are positive - need some kind of energy-input to keep them going. In the case of feedback in a sound-amplification system, where sound produced by a loudspeaker is fed back into the microphone and thence again to the loudspeaker via the amplifier, the energy input comes from the main power supply – without it the feedback loop will not sustain itself. The same applies in the case of backwash - instructors need some kind of motivation above and beyond the 'satisfaction of a job well done', to maximise their pass rates. This is provided by the 'performativity' paradigm of educational management now prevalent in many countries [3], [15], [16], and [17] in which target metrics (such as pass rates and distinction rates) have progressively replaced professionalism, and increasingly become goals in their own right rather than being simply measurements. The problem with this situation is that teachers under pressure, and hence forced – or, at least 'motivated' - to 'optimise' the efficacy of their teaching, might well be strongly inclined to 'teach to the test' in order to achieve 'acceptable' targets. For instance, good pass rates can be achieved by teaching a repertoire of routine operations and not much else.

Unfortunately, there is quite a lot else that should be taught and learnt and that we can ill afford to jettison from our de-facto curricula, for example – as we have here argued - problem-solving. If we exclude true problem-solving from our examination system, we are in effect excluding it from the curriculum, regardless of what the 'lip service' utterances might say. This may be a significant but under-recognised ingredient in the education 'crisis' from which our country – and perhaps others - is currently said to be suffering.

References:

[1] Jonassen D H 2000 Toward a design theory of problem solving. Vol. 48, *Educational Technology Research and Development*. pp 63–85.

[2] Tuminaro J and Redish E F 2003 Understanding Students' Poor Performance on Mathematical Problem Solving In Physics. *Phys Educ Res Conf*. 720: pp113–6.

[3] Biggs J 1996 Enhancing Teaching through Constructive Alignment. *High Educ* [Internet]. 32(3): pp 347–64. http://www.jstor.org/stable/3448076

[4] Cheng L and Curtis A 2003 Washback or Backwash: A review of the impact of testing on teaching and learning. In: *Washback in Language Testing.* Routledge. pp 3–17.

[5] Heylighen F 2015 Stigmergy as a Universal Coordination Mechanism : components , varieties and applications. In: Lewis, Ted & Marsh, Leslie (eds), *Human Stigmergy: Theoretical Developments and New Applications (Studies in Applied Philosophy, Epistemology and Rational Ethics).* pp 1–43.

[6] Bodner G M and McMillen T L B 1986 Cognitive restructuring as an early stage in problem solving. *J Res Sci Teach.* 23: pp 727–37.

[7] Bodner G M and Herron J D 2002 Problem Solving in Chemistry. In: Gilbert JK, De Jong O, Treagust DF, Van Driel JH, editors. *Chemical Education: Research-based Prcatice.* Dordecht: Kluwer Academic Publishers.

[8] Martinez ME. What Is Problem Solving ? *Phi Delta Kappan*. 1998;79(8): pp 605–9.

[9] Bodner G M and Domin D S 2000 Mental Models: The Role of Representations in Problem Solving in Chemistry. *Univ Chem Educ*. 4: pp 24–30.

[10] Schuster D, Undreiu A, Adams B, Hsu L, Henderson C, and McCullough L 2007 Multiple Modes of Reasoning in Physics Problem Solving, with Implications for Instruction. *AIP Conf Proc* [Internet]. pp184–7. http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.2820928

[11] Smith M *Toward a unified theory of problem solving: Views from the content domains* [Internet]. 1991 http://books.google.com/books?hl=en&lr=&id=WAhO5VObl5AC&oi=fnd&pg=PP3&dq=Toward+a+Unified+Theory+of+Problem+Solving:+Views+From+the+Content+Domain+,&ots=xowPooLLUj&sig=Wm1NxbqmN0a0HCJrLkBu-0YV-bw

[12] Gil-Perez D, Dumas-Carré A, Caillot M and Martinez-Torregrosa J 1990 Paper and Pencil Problem Solving in the Physical Sciences as a Research Activity. *Stud Sci Educ*. 18(1): pp 137–51.

[13] Garrett R M 1986 Problem-solving in Science Education. *Stud Sci Educ*. 13: pp 70–95.

[14] Clerk D, Naidoo D and Shrivastava S 2011 Contrasts between student and examiner perceptions of the nature of assessment tasks. In: *Book of Abstracts, 57th Annual Conference of the South African Institute of Physics.* 85.

[15] Ball SJ 2003 The teacher's soul and the terrors of performativity. *J Educ Policy* [Internet]. 18(2): pp 215–28. http://www.tandfonline.com/doi/abs/10.1080/0268093022000043065

[16] Ball S J 2012 Performativity , Commodification and Commitment : An I-Spy Guide to the Neoliberal University. *Br J Educ Stud* [Internet]. 60(March): pp 17–28. http://www.tandfonline.com/loi/rbje20

[17] Burnard P, White J. Creativity and performativity: counterpoints in British and Australian education. *Br Educ Res J* [Internet]. 2008 Oct [cited 2014 May 30];34(5): pp 667–82. http://doi.wiley.com/10.1080/01411920802224238

[18] Clerk D P P 2018 Stigmergic inhibition of the teaching and learning of heuristic problem solving in physics. Unpublished PhD thesis submitted to the University of the Witwatersrand, Johannesburg, South Africa.