A study of creative reasoning opportunities in assessments in undergraduate calculus courses

Ciarán Mac an Bhaird, Brien C. Nolan, Ann O'Shea & Kirsten Pfeiffer

Published online: 05 Jun 2017.

Submit your article to this journal ⌖

Article views: 225

View related articles ⌖

View Crossmark data ⌖

Citing articles: 1 View citing articles ⌖

Routledge
Taylor & Francis Group

Check for updates

A study of creative reasoning opportunities in assessments in undergraduate calculus courses

Ciarán Mac an Bhaird[a], Brien C. Nolan[b], Ann O'Shea[a] and Kirsten Pfeiffer[c]

[a]Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland; [b]CASTeL, School of Mathematical Sciences, Dublin City University, Dublin, Ireland; [c]School of Mathematics and Statistics and Applied Mathematics, NUI Galway, Galway, Ireland

ABSTRACT

In this article we present the findings of a research study which investigated the opportunities for creative reasoning (CR) made available to first year undergraduate students in assessments. We compared three first year calculus courses across two Irish universities using Lithner's framework. This framework sets apart imitative reasoning (IR) (analogous to rote learning and mimicry of algorithms) and CR (which includes plausible mathematically-founded arguments). We report on the differences between reasoning opportunities in specialist and non-specialist courses and compare our findings with those of other similar projects. In addition, we will provide a critical reflection on Lithner's framework as a method for classifying tasks in calculus courses and suggest a modification.

Introduction

This article reports on an investigation of the type of reasoning opportunities afforded to first year undergraduate students in a range of calculus modules. In particular, we look at differences in reasoning opportunities offered to specialist and non-specialist students. In Ireland (the setting for our study), as well as in other countries, mathematics courses for specialist and non-specialist students are usually distinguished by the level of complexity of the mathematics encountered, and by the degree of rigour of the mathematical approach. Neither of these implies that the quantity of reasoning opportunities should necessarily be different: we address this question here. We carry out our analysis by studying tasks from continuous assessment assignments and examinations using the reasoning framework developed by Lithner (2008). Lithner defined reasoning as "the line of thought adopted to produce assertions and reach conclusions in task-solving" (2008, p. 257); this is not restricted to formal proof and can include both high and low quality arguments. The framework classifies tasks as requiring imitative reasoning (IR) or creative reasoning (CR). In this article, we will use the word *task* to mean any piece of student work including homework assignments, tests, and final examinations.

It is a widely-held view that the study of mathematics promotes the development of thinking skills. However, there is a growing concern that students "can pass courses via

---

mimicry and symbol manipulation" (Fukawa-Connelly, 2005, p. 33) and, as such, may not develop conceptual understanding or problem-solving skills. The type and content of coursework issued plays a key role in student progress. A considerable amount of research (for example Boesen, Lithner, and Palm (2010)) demonstrates that the types of tasks assigned to students can affect their learning. Bergqvist and Lithner (2012) found that students are more likely to use rote-learning if they are exposed repeatedly to tasks which require lower levels of cognitive demand. As a result, students may become unable to solve unfamiliar problems or to adequately transfer mathematical knowledge to other areas. In contrast, Jonsson, Norqvist, Liljekvist, and Lithner (2014) showed that an emphasis on CR tasks could lead to higher levels of mathematical competence. Thus, tasks can be seen as student learning opportunities (White & Mesa, 2014).

It is interesting then to ask what kinds of tasks are usually assigned in introductory courses at university. Indeed Pointon and Sangwin (2003, pp. 671–672) emphasise that "Since assessment is such a powerful tool for influencing students' choice of learning style, a subject specific understanding of assessment tasks appears to be a pre-requisite for any serious educational dialogue." They developed a taxonomy which they used to classify course-work and examination questions. They considered 486 questions in total from two first year undergraduate mathematics courses at UK universities. They concluded that 84.5% of assignment questions and 71.2% of examination questions required the use of routine procedures or the reproduction of previously seen material.

In Sweden, Bergqvist (2007) analysed 16 examinations from introductory calculus courses at four universities. Using Lithner's framework, she found that 70% of the examination questions could be solved using IR alone and that 15 of the 16 examinations could be passed without using CR. Tallman, Carlson, Bressoud, and Pearson (2016) developed a new classification method called the Exam Characterization Framework and used it to classify 150 Calculus 1 examinations from a large range of university level institutions in the US. They concluded that the questions on these examinations mostly required a low level of cognitive demand, displayed a low level of real-life contexts, and rarely asked for explanations or for a demonstration of understanding of core concepts. They found that 85% of the tasks classified required students to either recall facts or apply procedures; they note however that their analysis did not factor in students' experiences in their calculus courses.

White and Mesa (2014) adapted an earlier version of the framework used in Tallman et al. (2016) and used it to study tasks (both graded and non-graded) from different sections taught by different instructors of a single calculus module. Their new framework concerns the *cognitive orientation* of a task and they summarised it by assigning their tasks to three broad categories: simple procedures; complex procedures; and rich tasks. They found differences in the cognitive orientation of tasks given by different instructors even if each used the same textbook; they also found differences between the types of tasks assigned as homework and those that appeared on examinations. Contrary to what might be expected, the examination tasks had a higher proportion of rich tasks than was the case for the homework tasks, although the majority of tasks in both cases were procedural. Darlington (2014) considered examination questions from quite a different type of course – that of first year Pure Mathematics modules in Analysis and Algebra in Oxford. She found that more than half of the marks available (54.1%) were assigned to questions

which asked students to apply conceptual knowledge to construct mathematical arguments.

Several considerations may explain the predominance of procedural questions on final examinations in Calculus courses. Tallman et al. (2016) report discrepancies between examination questions and instructors' perceptions of the types of questions that they set, with instructors overestimating the amount of high level questions on their papers. Bergqvist (2012) considered university teachers' perspectives on the reasoning required in calculus examinations and concluded that lecturers felt that having too many CR level questions would lead to high failure rates. Bergqvist (2005) suggested that some teachers believed that higher level reasoning was only relevant to high-performing students as these were the only students good enough to use and benefit from such reasoning. Iannone and Simpson (2015) considered 148 undergraduate mathematics students' opinions of assessment methods across two universities in the UK. They found that these students tended to have a preference for "traditional" assessment techniques rather than new innovative ones and that " … these students are most interested in being accurately assessed according to their abilities and … have deep-seated concerns about the unfairness of achieving marks through shared endeavour, with luck, without effort or without ability" (2015, p. 1063).

Ellis, Hanson, Nuñez, and Rasmussen (2015) highlighted the importance of appropriate assignments as a means which "provide[s] students with the practice needed for developing understanding, but leaves much more of the responsibility of the construction of knowledge to the students" (p. 270). Mesa (2010) focused specifically on the types of strategy and arguments evident in examples (and related text and solutions) in textbooks for both introductory and honours Calculus courses. She investigated the control structures used, that is, how the textbooks explained: (1) how to decide what to do; (2) how to determine that the answer has been found; and (3) how to establish that the answer is correct (Mesa, 2010). She found that in general, both textbook types gave explicit information on how to solve a problem and recognise an answer, however the introductory textbooks had less discussion of verification of correctness than the honours textbooks.

For our research, we considered many of the frameworks mentioned above when deciding how to classify and compare tasks in undergraduate calculus modules. We were interested in the kinds of opportunities for genuine mathematical activity that are available in first year undergraduate modules. We decided to focus on studying the kinds of arguments that might be expected of students and so Lithner's description of reasoning and method of classifying reasoning opportunities seemed appropriate for this study. Given that "reasoning" is not limited to formal proof, we felt that the reasoning framework is useful in studying the types of tasks usual in calculus courses, where students are expected to make plausible arguments and conclusions but rigorous proofs are often not required. Because Lithner's definition of reasoning can cover both high and low level arguments, and since constructing arguments is something that students in any university calculus module should be doing (Larsen, Glover, & Melhuish, 2015), the reasoning framework allows for the comparison of rigorous and non-rigorous modules in a meaningful way. This type of comparison is an important part of our study.

In White and Mesa's (2014) framework, a degree of interpretation is required by the researchers to categorise tasks. For example, it could be argued that their example of an "Understand" task – to state in everyday language a mathematical representation of a

rate of change – could be solved by the application of memorisation (and so become a "Remember" task). That is, the written prompts and symbols in the task may lead a student to provide a response without having "[determined] the *meaning* of the instructional messages" (White & Mesa, 2014, p. 679, emphasis added). In this framework, the role of the researchers as experienced instructors is crucial, as it allows and requires them to apply their familiarity with, e.g. "tasks that experience has told [them] were frequently proceduralized" (White & Mesa, 2014, p. 679).

Our application of Lithner's framework attempts to circumvent this subjectivity (although it is not entirely absent; see below) by emphasising the number of previous occurrences of tasks in module materials in the classification scheme.

Our research questions are:

1. What kinds of reasoning opportunities are available to students in first year calculus modules in Ireland?
2. What differences exist between the reasoning opportunities available in specialist and non-specialist modules?

We also discuss our use of Lithner's further sub-classification, our experience of using the framework and our observations on how it could be refined to take account of more complex IR tasks.

Conceptual framework

Lithner (2008) distinguishes between two types of reasoning: IR and CR. IR is further separated into: memorised (MR) and algorithmic (AR). Reasoning which is classified as MR should be of the following nature:

1. The strategy choice is founded on recalling a complete answer.
2. The strategy implementation consists only of writing it down. (Lithner, 2008, p. 258)

At undergraduate level, MR is seen most often when students are asked in assessment to recall a definition or to state and prove a specific theorem. AR is identified when:

1. The strategy choice is to recall a solution algorithm. [...]
2. The remaining reasoning parts of the strategy implementation are trivial for the reasoner, only a careless mistake can prevent an answer from being reached. (Lithner, 2008, p. 259)

Lithner labels a reasoning sequence as CR if it has the following three properties:

1. Novelty. A new (to the reasoner) reasoning sequence is created, or a forgotten one is re-created.
2. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.
3. Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning. (Lithner, 2008, p. 266)

CR can be separated into two subcategories: Local creative reasoning (LCR); and Global creative reasoning (GCR). An LCR-task is solvable using an algorithm but the student needs to modify the algorithm locally in a non-'trivial' way. A GCR-task does not have a solution that is based on an algorithm and requires CR throughout (Bergqvist, 2007).

In this study, we relate mathematical tasks to the type of reasoning they require. Using Lithner's criteria and based on the provided module material (i.e. lecture notes, textbooks, assigned tasks) we categorise the tasks as MR, AR, LCR or GCR tasks.

Methodology

The modules

We selected three first year undergraduate calculus courses from two Irish universities. We selected modules which would span the range of such courses on offer in Irish universities (from the more rigorous to the more computational) and for which we had access to full sets of notes and assessments. The three courses were: a business mathematics module, a module for science students, and a module for pure mathematics students. We will refer to the students in the pure mathematics modules as specialists (as they intend to complete a degree in mathematics) and to the other students as non-specialists (as they intend to complete degrees in other subjects). The students in the specialist modules usually have a stronger mathematical background than those in the non-specialist modules, and this is reflected in the differences in the minimum entry requirements of these courses. Entry to university in Ireland is based on the results of a high-stakes examination called the Leaving Certificate. Students in the pure mathematics module studied here are required to enter university with at least 110 points (out of a possible 125) in Mathematics, while students in the other two modules need to obtain at least 20 points in Mathematics. In practice, the majority of students in the non-specialist courses have higher mathematics scores than these minimum requirements. To elucidate the context, we give a brief outline of the mathematics encountered by students taking the different modules.

Module A: Business Mathematics
The focus of the content of this module is on applications of linear and low-degree polynomial functions in economics and business. The applications are described to students, who are expected to carry out prescribed routines and to supply interpretations of their results. Students are also required to construct and interpret graphs of functions of the types mentioned above, to study and apply the compound interest formula, and are required to calculate partial derivatives of low-degree polynomial functions of two variables. The emphasis is on learning to use and interpret mathematical procedures that have applications in the principal areas of the students' studies (economics and business). Students completed five online homework sets, each worth 3% of the total module mark. Each set comprised between six and 10 short-answer questions.

Module B: Science Mathematics
This module aims to develop an appreciation and understanding of Differential Calculus. The course started with a short introduction to functions in general and specific families of functions such as polynomials, rational functions, exponential and logarithm functions,

and trigonometric functions. The focus then turned to understanding limits and continuity (without formal definitions or proofs), and to the use of these concepts to sketch graphs and to motivate the calculation of rates of change. Students saw how to establish and use derivative rules. The course finished with applications of differentiation. There were seven written assignments worth, in total, 15% of the module grade, and each assignment required that students submit between six and eight questions.

Module C: Pure Mathematics (Integration)

The purpose of this module was to familiarise students with the concept of the Riemann Integral and introduce them to the standard tools for the calculation of integrals. Definitions of the Riemann Integral were given for step functions and later for functions in general. This was followed by a discussion about which functions are Riemann-integrable, the Fundamental Theorem of Integration and the application of the integral for calculating the area underneath a curve. This part of the course had a mathematically rigorous emphasis and the lecturer proved many results. The rest of the course was dedicated to techniques of integration, as well as applications. The course concluded with a brief introduction to numerical integration. The continuous assessment consisted of five assignments worth a total of 30%. Each assignment required that students submit between 10 and 12 questions.

Of the three modules described above, Module C was the most rigorous, which can be seen from the frequent use of formal definitions and proofs. The other two modules focused on developing students' intuition and on calculation. It could be said that Module A was less mathematically demanding than Module B; for example, the course concentrated on low-degree polynomials as opposed to the wider range of functions in the Science module, and the emphasis was on using procedures while there was more time spent in Module B on providing a mathematical validation of these procedures. This would be normal for these types of modules in the Irish university system.

Task classification

There were four types of data in this study: lecture notes, assignments, end of module examination questions, and for two courses, recommended textbooks. All data was assembled with the cooperation of the module lecturers. In order to experience and familiarise ourselves with the classification process as described in the previous section, we first selected and classified a sample of exercises from a textbook. We used the procedures presented by Lithner (2008) and Bergqvist (2007); the suggested method is that coders first construct a solution to the task, then compare it to the course notes and textbook examples, and make a judgement as to whether the task could be solved by IR, or whether CR is needed. The tasks are further classified into the MR/AR/LCR/GCR sub-categories using the descriptions provided by Lithner (2008). Note that, in a manner similar to that employed by White and Mesa (2014), the tasks are classified using only the hypothetical arguments constructed by coders and the information arising from course notes and textbooks, and not by using any real arguments constructed by students. We classified the sample tasks independently and then met to finalise our procedures. This process allowed us to discuss and agree on our classification methods for the module material.

The most challenging part of the process concerned our consistent classification of tasks into either the AR or LCR categories, or the LCR or GCR categories. After reviewing our initial coding, we decided to adjust the framework as follows: Classify a task as an LCR-task if the solution was based on an algorithm and students had to modify one sub-procedure; Classify a task as a GCR-task if two or more sub-procedures were new, if a proof aspect was the novel element, or if mathematical modelling was the novel element.

We then applied Lithner's framework to the assignment (both submitted questions and practice or tutorial questions) and examination tasks for each module. Each module was classified independently by two authors, neither of whom worked in the home university of the module. This inter-rating approach ensures reliability of the analysis of the course material from the different modules (see e.g. Cohen, Manion, & Morrison, 2000). We classified 632 tasks in total; examples of these classifications are illustrated below and further examples are available in Macan Bhaird, Nolan, O'Shea, and Pfeiffer (2014).

*Example 1 - **Module A, Task 6.1.2*** Determine the APR of an account if the nominal rate of 9% is compounded continuously. How long will it take an amount of money invested in this account to double in value?

This exam question was classified as **IR**, specifically AR. Similar examples and questions were covered in lectures, available in the textbook and had been asked on homeworks. Thus, students may solve the task by recalling and implementing a familiar algorithm.

*Example 2 - **Module B, Task 3.6*** Solve for x: $(x + 3)^{4/3} = 16$.

This practice question was classified as **CR**, specifically LCR. Students had not seen an example displaying all of the mathematical features of this task in either the lectures or textbook. The ideas for answering the question were covered in both, however students had not seen methods for solving equations involving non-integer powers of linear algebraic expressions, so this was regarded as new element.

*Example 3 - **Module C, Task 3.1.1*** Let f and g be integrable functions over the interval [a,b] such that $f \geq g \geq 0$. Furthermore, let A denote the area of the region bounded by the graphs of f and g between x = a and x = b. Explain why

$$A = \int_a^b f(x) - g(x)\mathrm{d}x.$$

This additional practice question was classified as **CR**, specifically GCR. Students had seen the definition of the area under the graph of a non-negative function, and had seen sample calculations. The task quoted requires an argument based on intrinsic mathematical properties: no algorithm is available to students to provide this argument.

The level of agreement between the raters was high: 98% for Business Mathematics; 91% for Science Mathematics; 71% for Pure Mathematics. These percentages refer to the classification of questions as IR or CR-tasks; the corresponding rates for the classification into the subcategories MR/AR/LCR/GCR were slightly lower (96% for Business Mathematics; 85% for Science Mathematics; 60% for Pure Mathematics). When we used Cohen's Kappa statistics to measure inter-rater reliability we found similar results: for the division into IR or CR tasks we found values of kappa of 0.957 (Business Mathematics), 0.767 (Science Mathematics) and 0.414 (Pure Mathematics). According to Landis and Koch (1977),

these correspond to outstanding, substantial and moderate levels of reliability respectively. Since the subcategories are ordinal, we used the intra-class correlation to measure reliability and obtained values of the ICC of 0.945 (Business Mathematics), 0.777 (Science Mathematics) and 0.569 (Pure Mathematics). These values correspond to excellent, good and fair levels of reliability (Cicchetti, 1994). One reason for the lower agreement in the Pure Mathematics course may be the higher proportion of CR tasks and the challenges involved in categorising such tasks. We report on results and challenges in the sections below. The researchers met and discussed items on which they differed to settle on agreed final classifications.

It should be noted that in this type of analysis, we are unaware of the individual students' backgrounds, motivation, engagement levels and especially their learning environments outside of the lecture hall – for example in tutorials, in Mathematics Learning Support Centres, etc. This is outside the scope of our study, and there are other works, e.g. Maciejewski and Merchant (2016), which consider students' study approaches and strategies and how these impact on their grades. In applying Lithner's classification method, we only use information gleaned from the textbooks and notes. This is a possible weakness with the study. However, this difficulty resonates with the position in which the lecturer finds themselves: they must make decisions on teaching and assessment in the absence of detailed knowledge of their students' prior learning experiences.

Results

We classified the tasks in the three courses under consideration using Lithner's framework. In the Business Mathematics and Pure Mathematics modules, the lecturers assigned practice questions (to be done in tutorials or in the students' own time), questions to be submitted (these counted toward the module mark), and examination questions. The Science Mathematics lecturer followed this pattern too but also provided the class with more challenging tasks (labelled optional questions). Tables 1–3 below show the results of our task classification for the three modules.

We found differences in the types of tasks used both between the three modules and between practice and summative assessment tasks within modules. Since our data is categorical and we wished to investigate the association between the reasoning levels and either the type of modules or the type of assessment, we used chi-squared tests, and when the numbers in the sub-categories were small we used the analogous Fisher Exact test (FET). There was a statistically significant difference between the proportion of questions classified at the CR level in the three courses ($X^2(2, N = 630) = 99.538, p < 0.001$). We saw a heavier emphasis on CR in the Pure Mathematics course, but in all three modules we

Table 1. Numbers of questions in each category for Business Mathematics.

Required Reasoning Type	Practice Questions	Submitted Questions	Exam Questions	Total
IR	*93*	*100*	*20*	*213*
MR	0	0	0	0
AR	93	100	20	213
CR	*55*	*2*	*2*	*59*
LCR	33	2	2	37
GCR	22	0	0	22
Total	*148*	*102*	*22*	*272*

Table 2. Numbers of questions in each category for Science Mathematics.

Required Reasoning Type	Practice Questions	Submitted Questions	Optional Questions	Exam Questions	Total
IR	*97*	*40*	*1*	*15*	*153*
MR	0	0	0	0	0
AR	97	40	1	15	153
CR	*19*	*15*	*27*	*2*	*63*
LCR	15	11	5	2	33
GCR	4	4	22	0	30
Total	*116*	*55*	*28*	*17*	*216*

saw a higher proportion of CR tasks in the practice questions and a smaller proportion on the summative assessments.

When we considered examination questions only (Table 4), we found a statistically significant difference in the proportion of questions classified in each subcategory between the Specialist (Pure Mathematics) and the Non-specialist (Business and Science Mathematics) modules (*FET, p* = 0.017); the specialist course was more likely to have more LCR tasks and fewer AR tasks than the non-specialist courses.

Table 5 shows the data for submitted questions and we see that there was a significant difference between the proportion of questions in each category for the Specialist and Service modules (*FET, p* < 0.001). The same was true also for the proportion of practice questions (*FET, p* < 0.001). In both cases, the relevant tasks in the specialist course were more likely to be classified as LCR or GCR than the tasks from the non-specialist courses.

Discussion

Reasoning opportunities in calculus courses

We have considered the reasoning opportunities made available to students in modules for specialists and non-specialists: it seems that most of the literature on the classification of the cognitive demand of university mathematics tasks concentrates on courses for one or other of these groups but not both. One of the benefits of Lithner's framework is that it is appropriate for use in both types of courses, allowing for comparisons between them.

Using the reasoning framework, we have been able to classify tasks in each of the three calculus modules in this study. We found that the Pure Mathematics Integration module had the highest proportion of CR questions for both practice tasks and assessment tasks (final examination as well as continuous assessment questions). The Business Mathematics module had a slightly higher percentage of practice tasks classified at the CR level than was the case in the Science Mathematics module, but the proportion of continuous assessment tasks at the CR level in the Business Mathematics module was very low. The fact that

Table 3. Numbers of questions in each category for Pure Mathematics.

Required Reasoning Type	Practice Questions	Submitted Questions	Exam Questions	Total
IR	*13*	*24*	*7*	*44*
MR	2	0	1	3
AR	11	24	6	41
CR	*51*	*43*	*4*	*98*
LCR	27	25	3	55
GCR	24	18	1	43
Total	*64*	*67*	*11*	*142*

Table 4. Classification of examination questions.

Course Type	MR	AR	LCR	GCR	Total
Non-Specialist	0	35 (89.7%)	4 (10.3%)	0	39 (100%)
Specialist	1 (9.1%)	6 (54.5%)	3 (27.3%)	1 (9.1%)	11 (100%)
Total	1 (2%)	41 (82%)	7 (14%)	1 (2%)	50 (100%)

the Pure Mathematics module had significantly more CR questions than the two non-specialist modules is unsurprising. The results found here for the service modules (i.e. 89.7% of examination questions were classified as AR-tasks) are very similar to those reported in the studies of examination questions in the US and Canada by Tallman et al. (2016) and Maciejewski and Merchant (2016), and slightly lower than those reported by Bergqvist (2007) for Swedish university examinations. The results appear significantly different to those reported in White and Mesa (2014). The proportion of examination tasks at the CR level in the Pure Mathematics course (36.4%) is lower than the corresponding results given by Darlington (2014), however the module in our study was a Calculus module and not an Analysis one, which may account for the different emphasis. MR-tasks were very rare in our data; we found none at all in the service modules and three in the module for specialists with one of these on the final examination.

Our analysis shows that the differences in proportions notwithstanding, all three lecturers made efforts to include CR tasks in their courses. This was achieved in different ways in each module. In the Business Mathematics course, the continuous assessment assignments were submitted through an online homework system and took the form of multiple choice short answer questions. The course instructor was able to choose the questions used but did not write new tasks, and this may account for the very low proportion of CR tasks in the marked assessments. However, the lecturer also assigned 148 practice questions, and 37% of these were classified at the CR level (with 15% at the GCR level and 22% at the LCR level). Thus, the students in this module were presented with high level reasoning opportunities in the coursework. The lecturer in the Science Mathematics module assigned two different types of practice questions; those designated as practice questions were mostly classified at the IR level (83.6%) but all but one of those designated as optional questions were CR tasks (96.4%), with 79% classified in the GCR category. It seems that both of these lecturers were aiming to strike a balance between the need for their students to practise basic calculus skills and giving them opportunities to work on more challenging problems. Ellis et al. (2015) identified the importance of high-level tasks (complex word problems questions requiring justifications or multiple representations …) in the homework assigned at institutions with successful calculus programmes; it is unclear whether the presence of more demanding tasks as practice questions on assignments has the same effect. The practice and optional tasks give students opportunities to engage in CR but if these questions are not submitted it is possible that students do not engage with them fully.

Table 5. Classification of submitted questions.

Course Type	MR	AR	LCR	GCR	Total
Non-Specialist	0	140 (89.2%)	13 (8.3%)	4 (2.5%)	157 (100%)
Specialist	0	24 (35.8%)	25 (37.3%)	18 (26.9%)	67 (100%)
Total	0	164 (73.2%)	38 (17%)	22 (9.8%)	224 (100%)

When we consider the finer classification into MR/AR/LCR/GCR tasks we see that there were very few GCR tasks in general, and especially on final examinations: there were none in either of the service modules and only one on the Pure Mathematics module. The numbers of GCR-tasks on submitted assignments was not much higher (none on the Business Mathematics module, four [7%] on the Science Mathematics Module, and 18 [27%] on the Pure Mathematics module). There were significantly more CR tasks on the Pure Mathematics examination than appeared on the service course examinations, however in all three modules, the proportion of CR questions on the final examination was much lower than the corresponding proportion for the tasks assigned during the course. This is to be expected, especially given the time constraints in a final examination. The low proportion of CR tasks may also be due to concerns about making the final examination too difficult (Bergqvist, 2012), and may relate to the thinking that results in different emphases in textbooks for specialist and non-specialist students (Mesa, 2010).

Reflection on applying Lithner's framework

We now discuss our experience of using Lithner's framework. We comment on the benefits and the challenges that this presented, and make some tentative proposals as to how the framework could be adapted to maximise these benefits.

Benefits of using the framework

Our experience of using Lithner's framework reinforced much of its original purpose in that it provided us with a tool for "use-inspired basic research" aimed at both "increased fundamental understanding and at contributing to developing teaching" (Lithner, 2008, p. 256). In particular, we found that it provided an empirically validated language for discussing key aspects of the teaching and learning of calculus. The framework reifies concepts of major concern to university mathematics departments (rote learning, teaching for understanding, teaching by, through and for problem-solving …) that are (in our experience) often discussed in less formal and less productive ways. It provided us with a tool to analyse and compare different calculus modules, and to re-think assessment strategies for those modules. The framework is well-adapted to the situation in which we find ourselves as mathematics lecturers: assessment tasks must be selected without having *detailed* knowledge about the students' prior learning experiences and engagement with different modes of reasoning. This framework and its categories provide a means of addressing this.

Challenges in working with the framework

Certain words used in the description of the categories of the framework required particularly careful interpretation. In describing Algorithmic Reasoning (AR), a key feature is that "the reasoning parts of the strategy implementation are *trivial* for the reasoner … " (Lithner, 2008, p. 259, emphasis added). The subjective nature of the adjective "trivial" is explicit in the qualifier "for the reasoner". We sought a degree of objectivity in rating the trivial/non-trivial character of different tasks. This was done on a consensus basis, where considerations such as our agreed opinion of the mathematical capacities of the median student in each given cohort were used. These were naturally different for each

of the three modules and were based on our experiences as mathematics educators in Irish universities and on the high degree of homogeneity of the prior mathematical experiences of those students. (A similar assessment influenced our classification of tasks as new or not new: see below.) Other factors also led to an unambiguous characterisation of a reasoning step as trivial, e.g. when that step required a near-identical reproduction of a text-book example (with, say, coefficients of a quadratic equation changed in a way that leaves the structure of the task-solution unaltered). This characterisation of *trivial* is provided in Lithner (2008, p. 264): nevertheless, we felt that a decision on how to deal with the (subjective) issue of triviality was required.

A similar situation arose in determining if a sub-procedure was "new" (Lithner, 2008, p. 266). This influences the categorisation of a task as requiring CR (and in distinguishing LCR/GCR). Our starting point was that an element of reasoning/strategy was considered *new* if it had not been used in the lecture notes or in the textbook. Occasionally this guideline did not suffice to make a decision: for example, if a previous question from the same or another set of homework tasks, for example, had required the same procedure. We agreed to categorise an element as *not new* if students had been required to complete it for summative assessment on a previous occasion, otherwise it was considered as *new*.

Many learners have difficulty with using their mathematical knowledge and skills in new contexts (Barnett & Ceci, 2002). Thus, we considered the question of context when rating reasoning as new or not new. If a task/procedure/sub-procedure required the application of mathematics in a context in which it had not previously been applied by the student, then it was seen as a new element. This could be a relatively simple step such as factorisation or finding a derivative, or demand a more advanced strategy or method. The key features leading to the categorisation of a (sub-)procedure as novel was that it was non-trivial (see above) and was not the central focus of the task at hand. For example, the application of a trigonometric identity in a definite integral was categorised as a new sub-procedure – provided this application had not previously arisen (in lecture notes, a text-book or a compulsory assessment task) in the same context.

The category *GCR* embraces a wide spread of reasoning demands. A sufficient condition for a task to be considered *GCR-task* is if a proof is required. While Lithner (2008) provides a commentary on the role of formal logical reasoning in his framework, we found instances where it was not easy to decide if a proving element was involved in a particular task. For example, the Pure Mathematics (Integration) module presented students with the following task:

> **Task 2.2(3):** *Determine the derivative of [the function]* $f(x) = x^t$ *for* $t \in R$ *and* x > 0. (Recall that $x^t = \exp(t. \ln(x))$ *and use the Chain Rule*).

Here students are required to apply the Chain Rule to develop a general rule. Is this just an application of the Chain Rule or should the fact that a general rule is being developed be seen as a proving element? We chose the latter, and considered this task as *GCR-task*. This is consistent with Lithner's categorisation of CR as involving new elements of mathematically grounded argumentation. This is consistent with the sequence of tasks in the relevant exercise set: the previous question asks students to determine the derivative of x^n, where n is a positive integer. Students are advised to "use induction and the product rule". Thus, it is made clear that argumentation based on intrinsic mathematical properties is required.

Is there room for an additional category?

We have highlighted that "the framework is intended for use-inspired basic research" (Lithner, 2008, p. 256), and on this basis, we tentatively suggest that the addition of a new category as described below may be of use to researchers applying the framework.

We have noted that a task was categorised as GCR (rather than LCR) task if several challenging sub-procedures are required – and that these sub-procedures meet the criteria of a CR task (novelty, plausible reason, anchoring to intrinsic mathematical properties). Lithner links such tasks to the use and development of important mathematical competencies by students (2008, pp. 269–270). The creative input is linked to the qualities of "fluency, flexibility and novelty" discussed by Silver (1997; cited in Lithner, 2008). These qualities come to the fore in tasks that avoid the reduction in complexity associated with IR tasks. However, in our research, we came across tasks which maintained a high degree of complexity, but which nevertheless required for their completion algorithmic (and hence imitative) reasoning only. Examples include certain integration tasks in the Pure Mathematics (Integration) module, for example:

Task 4.5 (4): Use an appropriate substitution to verify the following integral (take the point of view that you don't know the answer, so differentiating the right-hand side […] is not an option).

$$\int x^2 \sqrt{a^2 + x^2}\, \mathrm{d}x = \frac{x}{8}(a^2 + 2x^2)\sqrt{a^2 + x^2} - \frac{a^4}{8}\ln\left|x + \sqrt{a^2 + x^2}\right| + c$$

Such tasks appeared to require "structural conceptions" on the part of the students, and to breach the "modest dimensions of human working memory" in order to be successfully solved (Lithner, 2008, p. 269). In other words, students must show a degree of understanding of the relevant intrinsic mathematical properties of the components of the task; they must display fluency in the execution of a variety of sub-procedures, and they must be flexible in their choices of solution algorithms for those sub-procedures. But it may also be the case that there are no sub-procedures in the task-solution that require the novel reasoning associated with CR tasks. Hence it may be appropriate to introduce a new (sub-)category that has the characteristics described above as its defining property – a non-trivial, complex synthesis is at work, that requires more than working memory and leverages students' fluency and flexibility. Such tasks possess many of the features that, in Lithner's framework, play a central role in developing valuable mathematical capacities. We acknowledge that we have no empirical evidence – based, for example, on observations of students working on tasks (Lithner, 2000) – that such a separate category can be identified, nor that such tasks do indeed engender positive mathematical capacities. But given our experience as researchers using the framework for basic research, we consider that the question of the existence of this separate category of algorithmic reasoning (AR) – tentatively titled Complex Algorithmic Reasoning – is worthy of further investigation.

This new category would mirror to some extent the "Complex Procedures" category of White and Mesa (2014): there is a shared feature of requiring multiple steps to solve the task. Their category contains just one type of task – Recognise and Apply Procedures – which they define as one where "students must recognise what knowledge or procedures to recall without being directly prompted. … Students may have to string together several procedures" (White & Mesa, 2014, p. 680). This category owes its existence to the difficulty

in categorising tasks on the boundary between the categories of simple procedures and rich tasks.

We see we a link with mathematical creativity in the following way. Veale (2012) emphasises the computational and algorithmic mechanisms that underpin linguistic creativity. He argues that the creative use of language (for example, in humour or in poetry) is founded on, inter alia, "a mastery of everyday language" (2012, p. 15), and on the finessed search of constrained conceptual state spaces (2012, Chapter 3). (A poet, for example, might begin with a state that includes both a blank page and his vivid memory of some daffodils, and searches his word-stock – and engages in other creative processes – to produce a reflection on this memory that meets the constraints [conventions] of lyric poetry written in iambic metre with a particular rhyme structure.) The computational perspective ascribes an important role to these two features: mastery of everyday language, and skilful, subtle searching. We see these as analogous to the concepts of fluency and flexibility mentioned above. This leads us to conclude that the Complex Algorithmic Reasoning category tentatively described above, in which fluency and flexibility must be leveraged, may have an important role in developing students' mathematical creativity. More generally, these tasks (and likewise the Complex Procedures of White and Mesa, 2014), may have a useful role to play in what is referred to in Sonnert and Sadler (2015) as *Ambitious Teaching*; this term was defined by Jackson, Garrison, Wilson, Gibbons, and Shahan (2013) as pedagogy that seeks to actively engage students in class and to promote mathematically-founded class discussions.

Conclusion

Iannone and Simpson note that " … there is strong evidence in the research literature (Scouller, 1998) that the higher the stakes, the more assessment influences students' approaches to learning" (2013, p. 19). Seeing that development of mathematical reasoning skills is an important objective of teaching of mathematics at all levels, in particular at university, including tasks that require such reasoning in summative assessment seems imperative. However, several studies of assessment tasks indicate that often a relatively small amount of these tasks require CR. Findings from our study, which is based on three distinct first year calculus courses, indicate that in Ireland the situation may be different for specialist and non-specialist modules. While we found that the specialist module offered a relatively high amount of summative assessment tasks requiring CR, the corresponding proportion for the two service modules may raise concern. The less mathematically complex non-specialist modules need not have a lower proportion of CR tasks: advanced mathematical thinking is not "tied to a particular level of mathematics" (Edwards, Dubinsky, & McDonald, 2005). For example, the compulsory homework tasks of these courses required mostly AR or LCR, and limited numbers of these tasks required GCR. (One reason for this may be the difficulty of including tasks which require GCR when using an online-homework service). Final examinations for these modules again included mostly AR-tasks and minor LCR-tasks: GCR was not required.

Lithner's framework for categorising types of mathematical reasoning and ultimately mathematical tasks turned out to be very beneficial, both from a researcher's as well as from a lecturer's perspective. After gaining experience during a "test run", we agreed on classification criteria that were designed to be clear and unambiguous. Occasionally a

discussion between researchers was needed to agree on the categorisation of certain tasks. Thus, for research purposes, involvement of at least two researchers to categorise the same tasks is advisable. The framework provides an efficient lens to compare mathematical modules, for example modules for specialists and non-specialists. As the classification is based primarily on the course material, the framework would also be suitable for international comparisons. Lecturers may also consider the framework a useful tool to reflect on their own assessment practice. As we have seen from the work of Tallman et al. (2016), lecturers may consider that they are assigning high-level assessment tasks when closer analysis reveals that they are not. An awareness of the proportion of tasks requiring CR in summative assessment may trigger an informed debate on what a reasonable proportion of such tasks should be, in both specialist and non-specialist courses. After all, if one of our aims when teaching mathematics is to develop students' reasoning skills, then we should give them ample opportunities to do this.

Funding

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637.

Bergqvist, E. (2005). How students verify conjectures: Teachers' expectations. *Journal of Mathematics Teacher Education*, *8*, 171–191.

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, *26*, 348–370.

Bergqvist, E. (2012). University mathematics teachers' views on the required reasoning in calculus exams. *The Mathematics Enthusiast*, *9*(3), 371–407.

Bergqvist, T., & Lithner, J. (2012). Mathematical reasoning in teachers' presentations. *The Journal of Mathematical Behavior*, *31*(2), 252–269.

Boesen, J., Lithner, J., & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics*, *75*, 89–105.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London: RoutledgeFalmer.

Darlington, E. (2014). Contrasts in mathematical challenges in A-level mathematics and further mathematics, and undergraduate mathematics examinations. *Teaching Mathematics and its Applications*, *33*, 213–229.

Edwards, B. S., Dubinsky, E., & McDonald, M. A. (2005). Advanced mathematical thinking. *Mathematical Thinking and Learning*, *7*, 15–25.

Ellis, J., Hanson, K., Nuñez, G., & Rasmussen, C. (2015). Beyond plug and chug: An analysis of calculus I homework. *International Journal of Research in Undergraduate Mathematics Education*, *1*, 268–287. doi:10.1007/s40753-015-0012-z

Fukawa-Connelly, T. (2005). Thoughts on learning advanced mathematics. *For the Learning of Mathematics*, *25*, 33–35.

Iannone, P., & Simpson, A. (2013). Students' perceptions of assessment in undergraduate mathematics. *Research in Mathematics Education*, *15*(1), 17–33.

Iannone, P., & Simpson, A. (2015). Students' preferences in undergraduate mathematics assessment. *Studies in Higher Education*, *40*(6), 1046–1067.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education*, *44*(4), 646–682.

Jonsson, B., Norqvist, M., Liljekvist, Y., & Lithner, J. (2014). Learning mathematics through algorithmic and creative reasoning. *The Journal of Mathematical Behavior*, *36*, 20–32.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Larsen, S., Glover, E., & Melhuish, K. (2015). Beyond good teaching: The benefits and challenges of implementing ambitious teaching. In D. Bressoud, V. Mesa, & C. Rasmussen (Eds.), *Insights and Recommendations from the MAA National Study of College Calculus* (pp. 93–105). Washington, DC: MAA Press.

Lithner, J. (2000). Mathematical reasoning in school tasks. *Educational Studies in Mathematics*, *41* (2), 165–190.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, *67*, 255–276.

Mac an Bhaird, C., Nolan, B., O'Shea, A., & Pfeiffer, K. (2014). *An analysis of the opportunities for creative reasoning in undergraduate Calculus courses*. Proceedings of the 2014 Science and Mathematics Education Conference (SMEC). Retrieved from https://dcu.ie/smec/2014/oral.shtml

Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *International Journal of Mathematical Education in Science and Technology*, *47*(3), 373–387.

Mesa, V. (2010). Strategies for controlling the work in mathematics textbooks for introductory calculus. *Research in Collegiate Mathematics Education*, *16*, 235–265.

Pointon, A., & Sangwin, C. (2003). An analysis of undergraduate core material in the light of hand-held computer algebra systems. *International Journal of Mathematical Education in Science and Technology*, *34*(5), 671–686.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*, 453–472.

Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM*, *29*(3), 75–80.

Sonnert, G., & Sadler, P. (2015). The impact of instructor and institutional factors on students' attitude. In D. M. Bressoud, V. Mesa, & C. L. Rasmussen (Eds.), *Insights and recommendations from the MAA National Study of College Calculus* (pp. 17–29). Washington, DC: Mathematical Association of America.

Tallman, M. A., Carlson, M. P., Bressoud, D. M., & Pearson, M. (2016). A characterization of Calculus I final exams in U.S. colleges and universities. *International Journal of Research in Undergraduate Mathematics Education*, *2*, 105–133.

Veale, T. (2012). *Exploding the creativity myth: The computational foundations of linguistic creativity*. London: Bloomsbury Academic.

White, N. J., & Mesa, V. (2014). Describing cognitive orientation of calculus I tasks across different types of coursework. *ZDM*, *46*(4), 675–690. doi:10.1007/s1185801405889