Not all equals are equal: Decoupling thinking processes and results in mathematical assessments

Stephen Woodcock
University of Technology Sydney
stephen.woodcock@uts.edu.au

Abstract

One of the greatest challenges in mathematics education is in fostering an understanding of what mathematicians would recognise as "mathematical thought." We seek to encourage students to develop the transferable skills of abstraction, problem generalization and scalability as opposed to simply answering the specific question posed.

This difference is perhaps best illustrated by the famous – but likely apocryphal – tale of Gauss's school days and his approach to summing all positive integers up to and including 100, rather than just summing each sequentially.

Especially with the rise of technology-enabled marking and results-focussed tutoring services, the onus is on the educator to develop new types of question which encourage and reward the development of mathematical processes and deprioritise results alone. Some initial work in this area is presented here.

Introduction

> *"There is no agreed upon definition of mathematics, but there is widespread agreement that the essence of mathematics is extension, generalization, and abstraction"* (Hamming 1985)

In the literature there are many more general definitions of what constitutes mathematics than this one from the American mathematician and computer scientist Richard Hamming. These tend to speak in broad terms, defining any task described in numerical terms or any form of quantitative problem solving as being mathematics. However, these less tightly defined ideas of what lies at the heart of mathematical thinking fail to capture *why* mathematics is a vital part of universal education and why its skills are transferrable to multiple aspects of everyday life (Rombeg and Kaput 1999). A student may solve a problem or perform a numerical routine without employing the kinds of thinking which would be recognised as mathematical in nature (at least according to Hamming's definition.)

For example, school students tend to learn how to calculate the areas of common shapes or volumes of solids. In all likelihood none of these students will ever need to, for example, calculate the area of a rhombus in their later lives or careers. As such, if students simply memorise that they must calculate half of the product of the two diagonal lengths to obtain the correct answer to such a question, they have acquired a "trick" which is of minimal value, beyond scoring marks on the school examination. If instead, they are encouraged to understand the underlying geometry and congruent subshapes, they may develop skills which are more widely applicable to a whole range of other problems.

A Famous Example

Possibly the clearest example of this distinction between solving problems numerically and employing mathematical thought – i.e. approaches which are generalizable and scalable – is the often-told story of Carl Friedrich Gauss.

It is said that, when in elementary school, Gauss's teacher attempted to give the students the seemingly arduous task of adding together all positive integers up to and including 100. The teacher was assuming that summing these sequentially – 1+2 =3, 3+3=6, 6+4=10, 10+5=15,…etc. – would take the students a great time to complete as they would need to perform 99 separate addition calculations.

Gauss, as the story goes, recognised the sum as being what we would now know as the partial sum of an arithmetic sequence. He reasoned that 1+100=101, 2+99=101,…,50+51=101 hence the total was equivalent to the sum of 50 pairs of numbers each summing 101, so was equal to 5050.

Working with Hamming's definition of what lies at the heart of mathematics, Gauss's logic is certainly extendable, generalizable and handles the problem in abstraction. If the teacher had asked him to sum only the even numbers, he would be able to produce 25 pairs each equal to 102 and hence a total of 2550. If the teacher had asked him to sum the first million integers, or all integers between 1000 and 10,000, the approach employed would be robust to any such modifications or generalizations.

A Problem for Assessment

If we apply this idea to the example from Gauss's schooldays, let us consider the example of four hypothetical students as shown in Table 1.

Student A	**Student B**
1+2=3, 3+3=6, 6+4=9, 9+5=13… …4760+100	1+2=3, 3+3=6, 6+4=10, 10+5=15… …4950+100
Final answer 4860	Final answer 5050
Student C	**Student D**
1+99=100, 2+98=100, 3+97=100. This gives 49 pairs each summing to 100, plus 100 itself	1+100=101, 2+99=101, 3+98=101. This gives 50 pairs each summing to 101
Final answer 5000	Final answer 5050

Table 1: The methods employed by four hypothetical students attempting to sum the first 100 positive integers and the values obtained.

In evaluating the students' performance on the task, most assessors would agree that Student A's answer should be marked the lowest. It not only fails to employ a scalable approach, but also produces an incorrect numerical answer. At the other extreme, Student D's approach (which is that attributed to Gauss) would surely be awarded full marks.

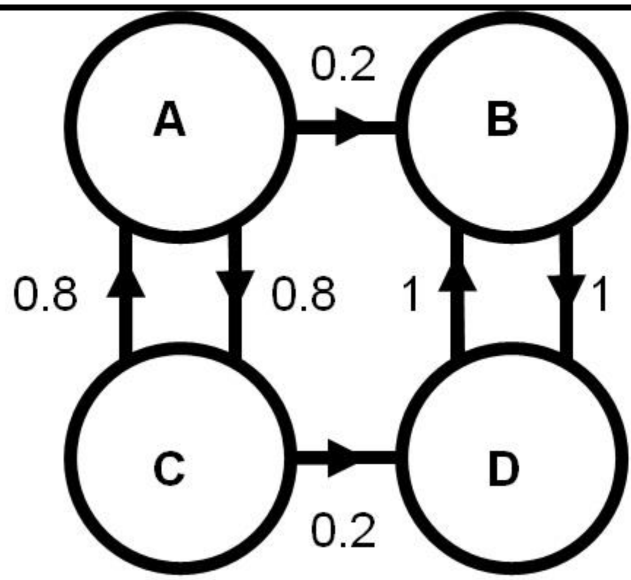
How to assess the other two students' performances is more problematic and would likely vary more widely between assessors and their marking criteria. Most mark schemes would suggest that Student C's approach would be worth some partial marks as the method employed was logically sound and mathematical in nature, but the final result was undermined by a careless oversight in missing that the number 50 did not belong to any of the 50 pairs which each summed to 100.It would, however be scalable to similar but larger problems and the student may be unlikely to make such an error again. By contrast Student B's approach would usually be awarded full marks as its final answer is correct, although its method is of little value to larger problems and shows minimal development of mathematical argument or generalizable skills.

Despite this deficit, it does not seem to be reasonable not to award Student B full marks as the student has answered the question correctly. What, then, are the alternatives to ensure that the marking criteria can separate students who can develop and display mathematical thought from those who can unthinkingly apply a brute force routine?

Traditionally, a question might have a "using method X" type statement, but this is clearly far from ideal; it takes away the student's opportunity to select the most reasonable approach for the question. Another option is to set problems with a sufficient time constraint such that a solution could only be reasonably be obtained if some degree of mathematical understanding were demonstrated. An example of this would asking a student to integrate an odd function on an interval which is symmetric about 0. For example, if asked to evaluate $\int_{-10}^{10} x^7 \cos(x)dx$, a student may obtain a correct solution via multiple implementations of integration by parts but this would be time-consuming. Another student may recognise that an odd function would necessarily integrate to zero on this interval. With sufficient time constraints, only one such approach would likely yield the correct answer within the allotted time. The downside of this approach, however, is that it may incentive students to simply guess answers. If there is a problematic integral which they feel they should be able to evaluate in under a minute, they may simply guess the answer without understanding the mathematical concept examined.

Example from Undergraduate Cohort

In 2018, a cohort of 98 students at the University of Technology Sydney enrolled in a first-year undergraduate level subject in Probability and Random Variables. When studying finite discrete Markov Chains and their equilibrium distributions, students had studied about both what these were conceptually and also that they could be obtained via eigenvalue-eigenvector methods from the transpose of the associated transition matrix. When presented with a tutorial problem (Figure 1) whose solution could be more easily obtained by simple inspection and understanding of the system, many students instead attempted to employ the much more arduous routine rather than the much more direct approach which they could have used if they had fully understood the problem.



Find the equilibrium distribution $\Pi_{eq} = \left(\pi_A \quad \pi_B \quad \pi_C \quad \pi_D\right)$ for the Markov Chain represented by the state transition diagram here where π_A is the equilibrium probability that the system is in State A etc.

Figure 1 The tutorial problem presented to students.

This problem can be solved, like all similar problems with finite discrete Markov Chains, via calculating an eigenvector of the transpose of the transition matrix, in this case $\begin{pmatrix} 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 1 \\ 0.8 & 0 & 0 & 0.2 \\ 0 & 1 & 0 & 0 \end{pmatrix}^t$. For this problem, however, it can be seen by inspection that the system will eventually reach a point whereby it is always in State B or State D, each 50% of the time hence $\Pi_{eq} = \left(0 \quad 0.5 \quad 0 \quad 0.5\right)$.

Student results were grouped based on each student's answer to this one tutorial question, by whether or not the student had attempted the more direct answer or not and whether or not the answer obtained was correct. After the (independent) final exam, the performance of all students within each of the four groups was recorded. Table 2 shows the pass rate for each group.

Method employed, Result	***n* =**	**Pass rate**
Direct, Correct	20	90%
Routine, Correct	24	75%
Direct, Incorrect	8	87.5%
Routine, Incorrect/No attempt	29	37.93%

Table 2: Pass rates for the four groups. From the cohort of 98 enrolled students 17 were absent either for the tutorial in question of the final exam and were excluded from this dataset.

Given the very small sample size here, few definitive conclusions can be drawn but this small study perhaps poses interesting questions for future larger evaluation projects. It is interesting to note that the pass rate for students who simply applied the more arduous eigenvector method and got the answer correct was lower than that for students who attempted to figure out the answer directly from the problem description.

Quasi-Multiple-Choice Questions

One possible approach to gain the benefits of problems which reward mathematical understanding as opposed to applying rote-learnt routines while not rewarding guesswork is quasi-multiple-choice questions. For such questions, students would be presented with a choice of which question to answer from a finite list and would then be required to provide an open-ended answer to their chosen problem. Each of the options would be chosen to be likely intractable in the relatively short time allocated if only routine tools were employed. One option would be set up so that it can be quickly answered by students who understand and observe that better methods may exist for problems of the type of that option. An example of this is given in Table 3

Evaluate **one** of the following definite integrals. Clearly state which part (a, b, c or d) you have answered.			
a) $\int_{-10}^{10} x^6 \cos(x) dx$	b) $\int_{-10}^{10} x^7 \cos(x) dx$	c) $\int_{0}^{10} x^6 \cos(x) dx$	d) $\int_{0}^{10} x^7 \cos(x) dx$

Table 3: An example of a quasi-multiple choice question.

All questions may be answered by multiple implementations of integration by parts. In all cases, however, this is a non-trivial task timewise. A student who understands that one of these problems involves integrating an odd function on an interval symmetric about 0 would readily choose to answer part b and give the answer 0.

At the moment, this quasi-multiple choice concept has not, to the best of my knowledge, been tested on a student cohort but may be implemented in the near future.

Discussion

Increasingly, student attainment is quantified and recorded at all levels from preschool right through to postgraduate education (Thomas and Klymchuk 2012). An unintended consequence of this – at least in mathematics education – has been the prioritisation of results (which can be readily measured) over skills and understandings (which are more difficult to quantify directly.) This is perhaps further worsened by the appealing rise of technology-enabled marking which, in general, will only check whether a result is numerically (or, in some cases, algebraically) equal to a known result and not how that result was obtained. Additionally, there is increasing evidence from around the world (Le *et al*. 2010, Trenholm *et al*. 2018) that many modern technology-dependent subject delivery modes are further creating opportunities for superficial learning at the expense of skill or thought development.

Much of the discussion must come back to a more philosophical question of why assessment is ever undertaken. Marks and grades should not be awarded as a prize for an answer or as a goal in and of themselves; they are supposedly a proxy for the skills a student should have demonstrated. Assessment questions which can be answered through rote learning or surface understanding may encourage students to maximise their total marks even if it this comes at the expense of genuine intellectual development. In recent years, there have been several interesting projects seeking to address such issues, perhaps most notably the Reframing Mathematical Futures II project, led out of the Royal Melbourne Institute of Technology (Siemon 2017). Focussing on the crucial Middle Years (Year 7 to Year 10), the project seeks to give *“teachers, textbook authors and curriculum writers a sense of what type of reasoning they can expect and encourage at each level and in what directions students’ reasoning should be developed”* (Stacey 2010)

The work discussed here seeks to extend and build on these initiatives in the context of fostering undergraduate-level mathematical thinking. Although ideas such as the quasi-multiple-choice question remain as yet untested, if properly refined and developed, they perhaps offer an opportunity to blend the best opportunities afforded by technology-aided assessment with the need to incentivise students to prioritise mathematical thinking processes as well as obtaining a numerically-correct result.

References

Hamming R (1985) Methods of Mathematics Applied to Calculus, Probability and Statistics *Prentice-Hall* ISBN: 0-13-578899-4

Le A, Joordens S, Chrysostomou S & Grinnell R (2010) Online lecture accessibility and its influence on performance in skills-based courses. *Computers & Education* 55(1):313–319.

Romberg T & Kaput J (1999) Mathematics worth teaching, mathematics worth understanding. In *Mathematics classrooms that promote understanding* (Eds: Fennema E & Romberg TA) *Routledge* ISBN: 978-0805830279. 3-17

Siemon S (2017) Developing Learning Progressions to Support Mathematical Reasoning in the Middle Years: Introducing the Reframing Mathematical Futures II Project.In *40 years on: We are still learning! Proceedings of the 40th Annual Conference of the Mathematics Education Research Group of Australasia* (Eds: Downton A, Livy S, & Hall J) *MERGA* ISBN: 9781920846305 647-650

Stacey, K. (2003). The need to increase attention to mathematical reasoning. In *Teaching mathematics in Australia: Results from the TIMMS 1999 video study* (Eds: Hollingsworth H, Lokan J, & McCrae B (Eds.) *ACER*. ISBN: 0 86431 641 0. 119-122

Thomas M & Klymchuk S (2012) The school–tertiary interface in mathematics: teaching style and assessment practice *Mathematics Education Research Journal* 24(3): 283-300

Trenholm S, Hajek B, Robinson C, Chinnappan M, Albrecht A & Ashman T (2018) Investigating undergraduate mathematics learners’ cognitive engagement with recorded lecture videos *International Journal of Mathematical Education in Science and Technology* 1-22