Review of the 2011 NCEA Mathematics Examinations

20/12/11

Summary

- Most of the seventeen mathematics examinations for 2011 reflect the relevant standards, are fairly written and presented and are free from error. Assuming that a good proportion of candidates are being awarded *achievement*, *merit*, or *excellence* then the teaching in our schools is also at a very good level.

- On the other hand, the number of questions in the 2011 examinations which warrant criticism is beyond an acceptable level and it is important for its credibility in the mathematics community that the NZQA acknowledges this fact and lets it be known that steps are being taken to correct the deficiencies.

- It should be noted that mathematics examinations are probably the most difficult to set without error, contradiction or confusion. Consequently, it may be appropriate to have different systems for administering them. Most of the shortcomings of the examinations reviewed could have been avoided by a rigorous checking process.

- Recommendation 1.
  Each examiner should have an associated co-examiner, *in addition to any who are involved in setting the paper,* whose sole purpose is to check that the paper reflects the associated standards and is free of errors and confusing wording.

- Recommendation 2.
  An additional person should check the examinations when they have reached what is hoped to be their final form. This person should be an expert in the associated area of mathematics.

- Recommendation 3.
  Examiners should be released from any obligation they may feel to set questions in an artificial context. They should use 'real world' contexts only when they are natural to the question or when the translation from such a context is the point of the question.

Review of the 2011 NCEA Mathematics Examinations

20/12/11

Purpose of the Review

In accordance with the terms of reference of this review, its purpose is to review the overall validity and reliability of the 2011 NCEA mathematics examinations. Clearly the degree to which the validity and reliability can be determined by studying the questions of the examinations, as opposed to doing calculations on measures of candidate performance on those examinations, is somewhat limited. However it is, of course, hoped and expected that an examiner will be able to produce a valid and reliable examination and it is possible, with experience, to make reasonable judgements as to validity and reliability. This review will study the questions and examination papers from the point of view of an experienced examiner. It will take the rather general view that validity is a measure of how well an examination is likely to reflect the standards and that reliability is a measure of how well the examination determines a candidate's ability in the sense that other valid and reliable examinations on the same standards and candidates would arrive at similar results. Clearly judgements about validity and reliability are, to some extent, subjective but, generally it is a matter of simple observation if a question is written in a confusing way, is ambiguous, or is simply incorrect. If a question has these faults, it must, at best, contribute a randomness to the examination outcomes which will reduce both validity and reliability. It would also disadvantage some candidates in a way which has no connection with the relevant achievement criteria; ie it would be unfair.

Under the heading, "Review of Examinations" this report will cover straightforward factual observations about the questions and will reserve other judgements for the "Conclusions" section.

Context

This review is to focus on five particular examinations but to comment on the *overall* validity and reliability it has been necessary to look in less detail at all the examinations. Consequently, it is possible to report that, in general, most papers do reflect the relevant standards, are fairly written and presented and are free from error. Assuming that a good proportion of candidates are being awarded *achievement*, *merit*, or *excellence* then the teaching in our schools is also at a very good level. The examinations to be reviewed in detail have, it is surmised, been chosen because they have attracted negative comment. Consequently, this review would look unduly negative if the previous two sentences are not kept in mind. It should also be kept in mind that the review is written without any access to what has been said about the papers by others except what appeared in a rather general newspaper column. However, in spite of the overall positive comments above, it must also be said also that the number of questions in the 2011 mathematics examinations which warrant criticism is beyond an acceptable level and it is important for its credibility in the mathematics community that the NZQA acknowledges this fact and lets it be known that steps are being taken to correct the deficiencies.

There are some questions throughout the examinations where the language used could be more rigorous and precise; many of these have not been singled out for comment because the language used in the questions is what is commonly used by students and probably by teachers also. Nevertheless, a high standard of precision in the language of the examination questions may help to encourage higher standard throughout the education system. Usually, precise language in mathematics helps understanding rather hindering it even though students habitually resist it.

Review of Examinations

Level 1 91028, Investigate Relationships between Equations, Table and Graphs

Question 1(a)&(b)
Good questions; require an appropriate balance of skills from the achievement, merit and excellence standards with the higher levels weighted towards the later parts of the questions.

Question 2
The presentation of information in this question is very confusing. Initially, there appears to be a conflict between the fact that "Ian works for 8 hours each day" and the table which gives payment for hours worked from 4 to 10. This initial confusion could have been avoided by stating that "Ian works for no more than 8 hours each day" and giving a table which applied for each day's work; ie 8 hours only and making it clear that less than 8 hours would apply for a partial day at the end of the job.

However, further problems arise when one tries to work out how much Ian charges for travel each day. A careful reading of the information given makes it clear that the payment P of the table does not include what is paid for the travel because Emma "*also* pays for Ian's travel to her home each day"; ie it is in addition to the payment P of the table. At this point part (b) of the question becomes impossible and candidates may well give up.

Part (c) of the question asks why "the graph rises more steeply after 8 hours". Ignoring the problems above and interpreting the question as the examiner probably intended, this is not true. The graph rises *equally* steeply in all regions except for a series of very short steeper steps at each multiple of 8 hours.

Question 3
This question contains four parts with a good mixture of skills required. It would be effective for discriminating between the merit and excellence candidates but has some straightforward parts for the others.

Level 2 90290, Solve Straightforward Problems Involving Arithmetic and Geometric Sequences

Question 1, parts (a), (b), and (c) and question 2 parts (a) and (b)(i) are all reasonable questions with a development from being straightforward to requiring more thought and interpretation although it is not clear which parts would require *excellence* skills. Perhaps question 2(b)(ii) was intended to fill this role.

Unfortunately question 2(b)(ii) can only be described as extremely puzzling in all ways. It follows the same scenario as the previous part questions which outline a situation which leads to the statement that "the number of new cases reported by all medical centres each day was 20% more than the number of new cases on the previous day." That is, the number of new cases is increasing daily. In 2(b)(ii) we are told that "the number of new cases reported each day continued to increase, until it reached a maximum" but, the next sentence, tells us that "after the maximum number of new cases reported had been recorded" the number of new cases could then be modelled by a formula which represents a daily *decrease from the time of the outbreak of the disease.* A clear contradiction or, at best, a weird piece of modelling. In either event candidates would not be able to make sense of the question and if any of them gave the required answer as to when the maximum number of new cases were reported it would be by pure chance. This reviewer has no idea of what was intended or required.

Level 2 90292, Solve Straightforward Trigonometric Equations

As with 20290, question 1 and question 2 parts (a) and (b) are all reasonable questions covering the three achievement criteria but, again, the last part, question 2(c) is impossible. The question implies that there is just one value of a for which the given equation has four solutions for x in the range $0 \leq x \leq 2\pi$. In fact, there two *ranges* of values of a with this property, not a single value. In any case, finding these ranges would be too difficult for all but a very few candidates. See appendix 1 for an attempt at a solution.

Level 3 90635, Differentiate Functions and Use Derivatives to Solve Problems

The questions of this examination are as would be expected as far as the skills and content to be covered are concerned but there are two criticisms to be made.

Question 1(d) asks for a derivative $\frac{dy}{dx}$ to be found "in terms of a and b". When the derivative is found, depending on the method used, it will be in terms of a, b, and y or possibly in terms of a, b, and x. It could even be in terms of a, x, and y or b, x, and y but cannot be written in terms of just the two constants a and b. The question does not ask for the derivative to be found "in terms of a and b *only*" so is not technically incorrect but it is misleading and may well cause worry and confusion.

The other worry about this paper is that two of the "problem" questions, 1(f) and 2(c), are questions which are common in many text books (perhaps with different numbers or in a slightly different form). They are probably intended to test some of the higher level skills (merit or excellence) but if candidates remembered doing questions very like them before (or seeing their teacher do them) then success would depend on recall of those solutions rather than the skills the questions are intended to test.

Level 3 90643, Solve Straightforward Problems Involving Probability

The examination for Level 3, 90643 differs from the others considered in detail in this review. The other four examinations appear to be appropriate as far as reflecting the standards and the curriculum is concerned and mainly the questions are good in clarity, lack of ambiguity and language used. Their problem is the presence of a 'rogue question' or two. This paper does not have questions which are contradictory, ambiguous or otherwise incorrect. However, there must be some concern about reflecting the standards and, for that matter, the title of the standard which includes the word *straightforward*.

Question 1(a) is easily done by using intuition as to the meaning of *independence* and the properties of mutually exclusive events but one point of systematising probability theory (ie making it a mathematical theory) is to avoid the pitfalls of using intuition. To answer the question rigorously requires the skills associated with excellence.

Question 1(b) requires the use of conditional probability which is mentioned in the explanatory notes associated with merit but not with achievement. It also requires either the use of the binomial distribution or the use of a quite difficult tree diagram. The binomial distribution is not specifically mentioned in the Achievement Standard document but tree diagrams are. In either case this is not an appropriate question for students who will gain achievement but not merit.

Question 1(c) would be quite difficult for any students who had not covered the binomial distribution so the same comments apply to this as for 1(b).

Questions 2(a) and 2(b) suffer from similar difficulties to 1(a) in that it requires a good understanding of the properties of conditional probability. Again, this is appropriate for merit but not for achievement.

Questions 2(c) is reasonable for achievement level as is question 3(a).

Question 3 is much more appropriate for this standard in that it has part (a) which is good for achievement level candidates and parts (b) and (c) which involve more understanding and sustained working.

Conclusions

Level 1 91028 This examination would have been a good one with every expectation that it would be both valid and reliable if it were not for question 2. The confusion would introduce a random element into the candidates responses (not good for reliability) and it would be unfair to those who never got to grips with it. The actual errors (assuming the examiner did not intend to say that the travel time was not included in *P*) would disadvantage those who read the question carefully and understood the actual wording; they would then have difficulty proceeding at all.

The shortcomings of this question would be understandable in a first draft but it is difficult to see how they got past even a moderately rigorous checking process.

Assuming the likely intention of the examiner, this question would have been a good one if it had been presented and explained clearly without errors.

Level 2 90290

This examination was satisfactory (with the slight doubt that it may not discriminate well at the excellence level) except for question 2(b)(ii). This was the last part question in the examination so it probably would not cause undue concern to most candidates ; they would simply avoid it. Provided candidates did not waste time trying to make sense of it, it is unlikely to have affected the validity and reliability unless it was the principal part of the examination which was intended to decide on excellence achievement.

On the other hand, it is inexplicable that the question was ever seriously considered for inclusion and it is hard to imagine what a model answer would have looked like. It should not have passed *any* checking process.

Level 2 90292

As with 90290, this examination is satisfactory for the most part and it is just the last part of the last question which is impossible to do. Again, it is hard to imagine what a model answer would have looked like or to imagine what may have been intended. It is so strange that there is nothing to be said about it other than that, it too, should not have passed the most rudimentary checking process. As indicated above, an analysis of the problem, which does not constitute a solution, is given in Appendix 1.

Level 3 90635

This examination had one part question with rather confusing instructions. These would probably affect the medium level students more than the others; the weaker ones would not notice and the best would be certain when they had a correct result. In any case, the confusion would not be helpful for the reliability of the examination and would be unfair to students who happened to be put off by it.

The examination also had the potential shortcoming that some candidates may well have previously seen questions which are very close to two of those in the paper. The solution of problems of this general type, which involves translating information from a described situation to a mathematical model, is something that many students find difficult. Consequently, the questions are probably intended to be used to help discriminate at the excellence level; candidates who had worked on them before would have an advantage which was not a consequence of their skill level. This would detract from the validity of the examination.

Level 3 90643

This examination has aspects which are different from the other four which have been reviewed. It does not appear to have questions which are incorrect or even confusing in their wording. However, it is likely that some questions, as asked, involve the candidates in much more difficult work than was intended. That is, the wording, although not confusing, may not have been carefully thought through. In any case, as it has been presented, the examination appears to be too difficult for most candidates and the questions are in an order which is unhelpful for the weaker candidates. Hence, it does not give the weaker candidates an opportunity to show what they *can* do and has an associated lack of validity and reliability. The reliance on an intuitive understanding of aspects of conditional probability and independence would disadvantage any candidates who, quite reasonably, expect to give rigorous answers based on the mathematical theory. Of course, the intuitive understanding is desirable to have but it should always be backed up by theory and proof.

Level 1 91037

Although it was not included in the scope of this review, the examination for this standard warrants some attention.

Recommendations

It should be noted that mathematics examinations are probably the most difficult to set without error, contradiction or confusion. Consequently, it may be appropriate to have different systems for administering them. Most of the shortcomings of the examinations reviewed could have been avoided by a rigorous checking process. Experience shows that many errors and/or confusions arise when a question is altered (often for good reason) after its original composition so particular care needs to be taken when alterations are made.

Recommendation 1.
Each examiner should have an associated co-examiner, *in addition to any who are involved in setting the papers,* whose sole purpose is to check that the papers reflect the associated standards and that they are free of errors and confusing wording. The examiner and this co-examiner should *independently* write out model answers each time a question is composed or altered. These answers should be compared; when differences occur, the source of confusion should be understood and eliminated. Perhaps a signing-off process could be in place to ensure that these steps are taken seriously.

Recommendation 2.
An additional person should check the examinations when they have reached what is hoped to be their final form. This person should be an expert in the associated area of mathematics. Perhaps two or three such people could cover all the examinations. These persons, too, should write out full solutions to compare with the earlier solutions.

Many of the questions in the mathematics examinations have questions which have been put into a rather artificial 'real world context'. It may be desirable to try to make mathematics 'relevant' in this way in the classroom but it is of no value in an examination where, because of the necessary artificiality and oversimplification it makes questions more difficult because candidates must get to grips which what can be an unrealistic context and translate between that context and mathematical theory. Setting questions in a context introduces a potential for misunderstanding, confusion and error. This is a likely source of the problems with examinations 91028 and 90290. Of course, for some questions having a context may be necessary or natural and in others, the translation process may be the point of the question.

Recommendation 3.
Examiners should be released from any obligation they may feel to set questions in an artificial context. They should use 'real world' contexts only when they are natural to the question or when the translation from such a context is the point of the question.

Appendix 1

Analysis of question 2(c) of examination 90292

The question states that the equation

$$\frac{6\sin^2 ax+7\sin ax+2}{3\sin ax+2}=0$$

has four solutions for $0 \leq x \leq 2\pi$.

It then asks candidates to find the value of a and solve the equation.

Solving the equation:

$$\begin{aligned}\frac{6\sin^2 ax+7\sin ax+2}{3\sin ax+2}=0 &\iff \frac{(3\sin ax+2)(2\sin ax+1)}{3\sin ax+2}=0\\ &\iff 2\sin ax+1=0\end{aligned}$$

provided $3\sin ax+2$ and $2\sin ax+1$ do not have any zeros in common, which they do not.

Hence

$$\begin{aligned}\frac{6\sin^2 ax+7\sin ax+2}{3\sin ax+2}=0 &\iff 2\sin ax+1=0\\ &\iff \sin ax=-\frac{1}{2}\\ &\iff ax=\frac{7\pi}{6}+2n\pi \quad \text{or} \quad \frac{11\pi}{6}+2n\pi\end{aligned}$$

where $n \in Z$, ie n is any integer.

Hence, the solutions for (a,x) are points on a collection of rectangular hyperbolae. If graphs of $ax=\frac{7\pi}{6}+2n\pi$ or $\frac{11\pi}{6}+2n\pi$ are drawn for $n=-3\ldots2$, then it is clear that there are an infinite number of solutions for a no matter what value a has but, if x is to lie between 0 and 2π then there are just four solutions if $\frac{23}{12} \leq a < \frac{31}{12}$ or if $\frac{-25}{12} < a \leq \frac{-17}{12}$. This could also be proved analytically.

Quite a nice exercise if you like that sort of thing but not appropriate for level 2!

Appendix 2

The Reviewer
As reviewer of the examinations above it seems appropriate to list my qualifications and experience.

I have a PhD in Mathematical Physics from the University [redacted]

I have taught mathematics at [redacted]

I have been a marker and often examiner for the Universities Entrance Board (a precursor to the NZQA) for at least three decades. This was sometimes in Mathematics and sometimes in Applied Mathematics at either Bursary or Scholarship level.

I have worked as a tutor for secondary school pupils (weak and strong) also over a period of at least three decades.

I understand both the difficulty and importance of making examinations as fair, valid, and reliable as possible.

Signed

[redacted] 20/12/2011