AN EVALUATION OF A PROTOTYPE FEEDBACK TOOL FOR TEACHERS

SUMMARY

This project explores the utility of an item-bank analysis interface as a feedback tool for teachers. One potential use of item-level data is to provide teachers with information on the relative expectations of AQA for performance on an item and the overall test. Being able to analyse data at question level can provide teachers with additional information, enabling them to compare the standard set in AQA assessments against the performance level achieved by their students. A prototype item-bank analysis interface was demonstrated to a focus group of six examiners in order to ascertain their views upon the following:

1. What are the key functions desired by teachers from an analysis tool?
   Eight key functions were identified from the focus group.
2. How would teachers like to use an analysis tool?
   Teachers would like to use an analysis tool as both a post-exam analysis tool and a diagnostic tool prior to exams.
3. Would the analysis tool pose a risk of teaching to the test?
   None of the participants felt that an analysis tool posed a real threat of encouraging teaching to the test, so long as it did not contain model answers.
4. What training support would teachers require if an analysis tool were marketed to centres?
   All participants felt that a training day and a training manual would be the most appropriate form of support AQA could provide to teachers.

Additionally, this project explored:

5. The design issues related to the prototype item-bank analysis interface.
   A diagnostic tool prior to submission of an on-demand exam would be useful. However, the participants were disinclined to use functions in the prototype when they could not understand how the statistics were calculated. This has great implications for the implementation of on-demand testing. If teachers do not understand the new system they may not have confidence in the new examinations.

BACKGROUND

This project explored some analysis options AQA can pursue in repackaging the items retired from an item-bank. Past paper analysis will be made redundant when on-demand testing is implemented in 2015. Item analysis tools could be provided to teachers to act as a replacement. A prototype item-bank analysis interface has been developed within the Research and Policy Analysis department (RPA). The prototype analyses students’ responses at individual item level as well as test level. Feedback on school, class and individual performance is provided by the prototype. This project aims to use the prototype item-bank analysis tool to investigate the utility of item analysis tools in providing feedback for teachers.
This project aims to answer the following questions:

1. What are the key functions desired by teachers from an analysis tool?
2. How would teachers like to use an analysis tool?
3. Would the analysis tool pose a risk of encouraging teaching to the test?
4. What training support would teachers require if an analysis tool were marketed to centres?
5. What is the utility of item-bank analysis tools in 2015?
6. What are the design issues related to the prototype item-bank analysis interface?

METHOD

A small focus group, comprised of six AQA examiners, evaluated a prototype of an item-bank analysis tool. The participants were all examiners of GCSE Mathematics B 43001, 43003 and 43005, foundation and higher tier. In addition, all the examiners had teaching experience; five examiners were currently teaching, and one was a retired teacher. Half of the participants had used test feedback software before. The prototype included data from the June 2008 Mathematics B 43001, 43003, and 43005 papers, at foundation and higher tier.

Participants were given a short introductory talk where they were informed of the background behind item analysis tools, the aims and objectives of the day and the structure of the study session. Participants were then given a demonstration of the prototype software. After the demonstration had concluded, participants were invited to analyse their own students' data. Only two examiners had students who sat the Mathematics B June 2008 exams. However, dummy data was provided for those who did not have their own students' data to look at. Following this, participants were invited to complete a questionnaire on the desired functions of analysis tools (see Appendix). Responses from the questionnaires were totalled in order to provide starting points for the group discussion that followed.

RESULTS

1. What are the key functions desired by teachers from an analysis tool?

Participants were presented with functions of the system on the questionnaire and were asked to rate how useful they perceived each function to be. Ratings were made on a scale of 0 to 3, where 0 was 'not useful' and 3 was 'very useful'. The ratings provided by the participants were rank ordered and the median rating determined for each function. Participants felt that all the features suggested on the questionnaire were useful (Table 1). In the discussion, participants commented that the key data they would like in a report is the name, gender, raw mark, grade (when one can be provided), and national average.
### Table 1. Median rating of key functions desired by teachers.

<table>
<thead>
<tr>
<th>Functions</th>
<th>Rating Scale = 0 Not Useful, 1 Limited Use, 2 Useful, 3 Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant</td>
</tr>
<tr>
<td></td>
<td>1  2  3  4  5  6</td>
</tr>
<tr>
<td>Viewing the written question</td>
<td>3  3  3  2  3  3</td>
</tr>
<tr>
<td>Tests online</td>
<td>3  3  3  3  3  3</td>
</tr>
<tr>
<td>Seeing the questions an individual student got right/wrong</td>
<td>2  3  2  2  3  3</td>
</tr>
<tr>
<td>Comparing Schools marks with national average</td>
<td>2  3  2  1  3  3</td>
</tr>
<tr>
<td>Viewing question difficulty</td>
<td>3  0  2  2  3  3</td>
</tr>
<tr>
<td>Viewing assessment objectives and learning outcome.</td>
<td>3  0  2  3  1  3</td>
</tr>
<tr>
<td>Comparing individual marks with expected mark for each grade</td>
<td>2  3  2  1  2  3</td>
</tr>
<tr>
<td>Viewing your school's responses for each question compared to the national average</td>
<td>2  2  2  1  3  3</td>
</tr>
<tr>
<td>Generating a report</td>
<td>3  1  2  1  2  3</td>
</tr>
<tr>
<td>Link to the school IMS available on the internet.</td>
<td>2  1  2  1  3  3</td>
</tr>
<tr>
<td>Available on the internet.</td>
<td>1  3  2  1  2  3</td>
</tr>
<tr>
<td>Compare class marks with the expected marks for each grade</td>
<td>2  2  2  0  2  3</td>
</tr>
<tr>
<td>Comparing class responses for each question compared to the school average</td>
<td>3  1  2  1  2  2</td>
</tr>
<tr>
<td>Comparing male and female marks</td>
<td>2  2  2  1  2  2</td>
</tr>
<tr>
<td>Viewing the percentage of students achieving each mark on each question</td>
<td>2  2  2  0  1  3</td>
</tr>
</tbody>
</table>

Analysis from the questionnaire revealed that the key desired functions were the ability to:

1. View the written question
2. Access tests online
3. See the questions an individual got right/wrong:
   "We could really analyse by individual their strengths and weaknesses... you could use it as a tool to sort of direct your teaching" - Participant
4. Compare school marks with expected marks for each grade
5. View the question difficulty
6. View the questions by assessment objectives and learning outcomes:
   "If you could view the questions by assessment objectives and learning outcome – I think that would be very useful. Because if you are using it diagnostically that's what you want" - Participant

Upon discussion, the following key desired functions were identified:

7. For mock papers to be provided:
   "What I found very difficult is finding decent preparation material in the way of practice papers ...I was wondering if someone could be detailed to cobble together papers as practice papers – you know, AQA's official practice papers with this analysis back up, that could be purchased, we would certainly be interested in that" - Participant. Others strongly agreed.
8. To construct and analyse custom tests, with items arranged by subject area:

   “Right I really haven’t got through to this lot how to do X Pythagoras’ say, you know, if I can round up out of all the past papers the all the Pythagoras questions, and see the range and think about them and take my students through them” - Participant

Upon discussion, it became apparent that the features the participants favoured were dictated by their centre's requirements. There was variation between participants in the functions that they thought would be useful to them. Taking into account the limitations of using a small focus group, a wider survey of teachers would provide a better indication as to the key functions which are desired nationally.

2. How would teachers like to use an analysis tool?

2.1. Post-exam vs. custom tests
Participants reported on the questionnaire that they were more likely to use an analysis tool for post-exam analysis than for creating custom tests for learning or revision purposes; however, during the discussion, participants did voice the desire to create custom tests.

   “I don’t think the functionality should be lost to make up your own papers as well…that’s still a valuable thing.” - Participant

While participants were informed of the function to construct custom tests, this was not included in the demonstration. Participants were unsure as to how this process occurred, which may account for the discrepancies between the questionnaire responses and the comments in the group discussion. In the discussion, participants commented that an analysis tool would be of benefit to use prior to exams.

   “In an ideal world I suppose you would do it with the actual GCSE exams, look back and see how the year 11’s did and how that would assist your planning for the next year’s teaching. But I think it is more likely that you would do it with students who were coming up to GCSE.” - Participant

It became apparent that some teachers were reluctant to analyse custom tests because of the time it takes to create the custom tests.

   “Although the problem with doing it yourself is it is time-consuming so therefore teachers would not be so prone to do that.” - Participant

Participants were unenthusiastic about incorporating their own items into a test, but still felt that it could be one possible way that they would use an analysis tool.

2.2. Sharing information
Participants reported that they would definitely use an analysis tool to provide their students with feedback, and share the information with colleagues in the subject area. Participants were disinclined to share their students’ information with the head teacher and the students’ parents; however, they still considered this to be a possibility. A few teachers commented that they were concerned about sharing information with the head of department, because it could lead to ‘micro management’ and criticism. However, not all of the participants shared these concerns and commented that it depends upon the centre you work in. Furthermore, it was commented that if you were to use an analysis tool diagnostically, then such issues were unlikely to be problematic.
3. Would the analysis tool pose a risk of encouraging teaching to the test?

Participants were asked to rate potential risks of using an analysis tool on a scale of 1 to 11, where 1 represented no risk and 11 represented great risk. Participants rated analysis tools as posing a moderate risk of increasing the amount of testing in schools (Figure 1). Participants reported that there was a low risk of analysis tools encouraging teaching students to the test, without developing pupils’ understanding of the subject.

![Figure 1. Graph to show rating of risk of using an analysis tool.](image)

Upon discussion, it was clear from the teachers’ comments that they did not perceive the analysis tools proposed by AQA as posing a risk of encouraging teaching to test. When asked whether an analysis tool would encourage teaching children to the test, the following comments were made;

“*It’s just past papers with a bit of statistical analysis really.*” - Participant

“That’s why you provide mark schemes, to help them understand… Without going through the mark scheme with them they don’t learn. So I think it’s teaching them properly.” - Participant

“If you have got through your syllabus in year 11, then you do do virtually nothing but old papers, don’t you. And you come across some questions, with or without this diagnostic tool, and you think ‘we haven’t really got this cracked’ you stop doing old papers for two or three lessons and you crack into Pythagoras or what-have-you. And I can’t see that a sensible teacher is going to do anything different and you can’t legislate for the others.” - Participant

“I think it’s when you start putting in model answers that people really get into trouble.” - Participant
4. What training support would teachers require if an analysis tool were marketed to centres?

All participants felt that a training day and a training manual would be the most appropriate form of support AQA could provide to teachers.

5. What is the utility of item-bank analysis tools in 2015?

The flexibility an item-bank analysis interface offers may mean that teachers could use it as a tool to evaluate their pupils prior to submission to an on-demand test. The feedback from the group discussion suggests that the participants would be interested in using an analysis tool as a diagnostic tool prior to exams. However, participants were alarmed when informed that an analysis tool could provide an estimated grade on a custom test. Participants felt that this was misleading and dangerous, they could not understand how a grade could be estimated on a paper that had not gone through an awarding process. The participants were mistrustful of the statistics produced in the prototype software when they could not understand how they were calculated. An explanation as to how an analysis tool analyses performance may need to be incorporated into a training day, but the legal implications of predicting grades may preclude use of this feature.

This has great implications for the implementation of on-demand testing. Due to the complex nature of item calibrations and item response theory, which underpin on demand testing, and the radical change to processes, teachers may not understand how AQA calculates students’ grades, of course this assumes that they know how AQA currently calculates students grades! If teachers do not understand the new system, they may not have confidence in the new examinations. The current study is based upon the views of examiners; teachers who are not involved in the current examination process may differ in their reaction to the new examinations.

6. What are the design issues related to the prototype item-bank analysis interface?

During the demonstrations, participants voiced a few criticisms of the prototype which may be worth noting for development of an analysis tool:

6.1. Graphs and tables must be clearly labelled
Within the prototype, some of the graph legends were not easy to interpret. Teachers commented that they would have preferred labels which did not use ‘jargon’ and were clear.

6.2. Would like to view item breakdown by question parts
Participants were frustrated that the item breakdown on the prototype would allow them to view the number of students achieving each mark in a several mark question, but did not allow them to see on which part of the question these marks were gained or lost.

6.3. Data entry kept to a minimum
Participants felt that the time required for using an analysis tool would ultimately determine whether they would use the program in their school. Concerns were raised as to how many administration staff would be required for data entry. Some felt that if an analysis tool were to require a lot of staff time then they would prefer to produce the essential information they need though a self-made spreadsheet.
“I could do a quick spreadsheet for myself and run it off and put ticks and crosses against each student’s name … and I would have a quick pattern there and I think I would probably do that as I was marking and I would do it in half the time of actually entering it in this program.”
- Participant

6.4. Scales which do not have an intuitive meaning should be left off
Participants did not like the ability score produced in the report the prototype generated, as the scale held no meaning for them. Additionally, they felt that telling a pupil that they had a -2.4 ability score was likely to be harmful. Other scales which did not hold meaningful values to the teachers were also criticised:

“Those difficulty scales… if they are arbitrary why have them there at all? You can have a scale and it’s clear that there are the easy ones and those are the hard ones, the numbers, if they don’t mean anything why put them in?” - Participant

6.5. Estimated grades should cater for foundation and higher tier
The prototype uses item response theory to calibrate students’ responses and their estimated grade. The graph it produces does not change for foundation and higher papers. The participants were alarmed to see that their foundation paper student appeared to achieve an A by the prototype’s calculation. They felt that this was misleading and would be dangerous to present this information to pupils and parents. An analysis tool should only display the possible grades achievable on a foundation or higher tier paper.

The focus group is a small sample of teachers and may not be representative of all teachers. The results of this evaluation should be treated with caution.

CONCLUSION

This report indicates that there is a demand for item analysis tools amongst teachers. Participants reported that they thought the analysis functions proposed were useful and that they would be likely to use it in their schools, providing that time and cost demands were not high. It would be beneficial to further investigate item analysis tools as a feedback tool for teachers.

The prototype is currently limited to paper based testing. It may be beneficial for a new prototype with online facilities to be developed. Online testing would eliminate the need for data entry, thus reducing the demand on teachers’ time. The online prototype should focus on the key functionality revealed by the current study and improve on the design issues highlighted. The online prototype could be evaluated by a more representative sample of teachers, providing a stronger indication of the utility of an online analysis tool.

Victoria Spalding
16th October 2008