"THEY'VE ALL GONE TO COLCHESTER ZOO"

STAKEHOLDER VIEWS ON THE CHALLENGES OF DELIVERING MORE FLEXIBLE TESTING FOR NATIONAL HIGH-STAKES EXAMINATIONS IN ENGLAND

Christopher Wheadon

Summary

This research investigated the case for more flexible testing for national high-stakes examinations in England through a series of focus groups with teachers, pupils and examiners. Sample sizes were generally small, and the focus was on the examination of GCSE Science in the state sector, so care should be taken in generalising the findings beyond this qualification and sector. The consensus appears to be that more dates for examinations throughout the school year would be welcome as they would allow teachers to plan more effectively and reduce stress on pupils. More flexibility would allow teachers to manage the different demands placed on different academic streams more effectively, and build in contingencies when pupils progress at a different rate than anticipated. There was no appetite whatsoever for flexibility at an individual pupil level, however, as this was felt to be a recipe for chaos. Rather than leading to no child being left behind it was felt that it would be impossible to timetable and teach, leaving pupils without adequate support. The key concern of teachers, which was not entirely shared by pupils and examiners, was that changes introduced to enhance flexibility could impact on the security of the examinations.

Introduction

Current methods of maintaining standards in national high-stakes examinations in England have been shown to be inflexible (Wheadon, 2009). Value-added based statistical approaches depend on large stable cohorts being tested at similar points in their programme of study, while the weak criterion referencing approach requires committees of experts to be convened every time a test has been sat (Baird, 2007). These restrictions are only an issue if more flexible testing is required. But why would it be? The system has served well since 1918, if not earlier (Tattersall, 2007). One driver for change is the move towards on-screen assessment. At present, the take-up of on-screen assessment is low as schools and colleges have a limited number of computer terminals that can be used for assessment at any one time. As there is only one version of any test available for a test session, security concerns require every test to begin and end on a set time on a set day. If multiple versions of tests were available then the testing sessions could be lengthened, boosting the take-up of on-screen assessment. The second driver for change is the policy agenda that would see learners with their own personal path through learning, accompanied by their own examination timetable (Gilbert, 2006). While this prospect may seem remote, the system is continually being stretched in that direction. Curriculum 2000 split A-levels into modules that can be taken over the course of two years. Both mathematics and science GCSEs now offer three testing sessions per year. From 2011, functional skills qualifications will be offered six

AQA

times per year. The standard-setting methodology is struggling to keep up with the pace of change.

Wheadon (2008) has suggested that IRT test-equating designs may solve the standard-setting problems of a more flexible examination timetable, however they require some changes to test delivery that may be contentious. The need for some overlap between the test-takers and the tests offers the assessment agency choices that may be more, or less, palatable to stakeholders. One question that arises, for example, is the extent to which pre-testing can be built into live testing to support a test-equating design. Would candidates mind? Would their teachers mind? Would their parents mind? It is also extremely challenging to make IRT test-equating designs as secure as the current system that keeps all test items secret until the day and time of the designated test, and then releases all those items for public scrutiny, never to be re-used. To what extent are stakeholders prepared for the security of testing to be sacrificed, in some degree, in exchange for more flexibility? This research sought to answer some of these questions to begin to understand what some key stakeholders in the system (teachers, pupils and examiners) value in the way that qualifications are currently delivered, and to investigate the benefits they perceive would accrue from a more flexible system. It may seem perverse to ask these questions of the stakeholders – patients would not be consulted in the process of drug testing – but the age of the unquestioned and unquestionable expert is gone. Where public confidence is grounded in false beliefs, those beliefs are likely to be challenged and trust threatened (Newton, 2005). In the competitive market for general qualifications, a loss of confidence could also be accompanied by a loss of market share.

Methodology

The aim of this research, therefore, was to explore opinions, attitudes and values of teachers, pupils and examiners regarding the flexibility of examination delivery. While other groups are obviously of interest – parents, governors and the regulator – to some extent these groups are all guided by the views and experience of teachers and pupils. It is possible to conceive of a matrix of qualification level, school sector and subject that should be sampled from to gain a view across qualification, sector and subject, but such a matrix is beyond the scope of this research. Instead the focus was placed on the examination of GCSE Science in the state sector. GCSE Science lends itself to a flexible, modular approach and attracts the largest entries from the state sector. Focus groups were chosen as the appropriate methodology as they allow for a full discussion through which individuals can debate issues and thereby achieve clarity on their own positions (Stewart & Shamdasani, 1990).

Participants

The learner focus groups

In total, sixteen learners participated in the learner focus groups, which were held on three separate occasions. The first group consisted of four first year university students who were asked to draw on their experiences from school. It was hoped that they could offer a mature, but recent, perspective on examinations in general. An opportunistic sample was taken from a local university without particular reference to the subjects they had studied or the type of school or college they had attended. This is the only group who were not asked to draw on experience of GCSE Science. The second and third learner groups consisted of six year 10

pupils from a selective state school, and six year 10 pupils from a non-selective state school, all studying for the modular GCSE Science syllabus.

The teacher focus groups

The teacher focus groups were held on two separate occasions: the first with three teachers from the GCSE Science department of a selective state school, the second with nine participants on a GCSE Science day. These nine participants all came from the state sector; their schools varied in size from a small state boarding school to the largest comprehensive school in the county with over 2,000 pupils. To supplement these groups, the Deputy Head from a Special school was interviewed separately.

The examiner focus group

Only three examiners were recruited for the examiner focus group: the three Principal Examiners with current and direct involvement in compiling the modular GCSE Science A question papers. This group was chosen as they had experience of working with a large number of different structures for science at this level, going back to the 1970s, and are currently responsible for producing the examination papers for the new syllabus.

Sample size

The last round of focus groups with the teachers and pupils revealed enough redundancy in the data to imply that the sample size, for GCSE Science at least, was adequate. For this reason, no more focus groups were held. In terms of generalising beyond GCSE Science, the pupils' views did not seem to relate to particular subjects or levels, however examiner and teacher views for other subjects and levels, and teachers' views from different sectors on certain aspects, could be quite different to those reported here.

Group dynamics

The group dynamic aspect of focus groups was problematic for the first teacher group as the presence of the Head of Department may have led to motivated responding. Given more time and resources, interviews may have led to more authentic responses. For pupils, the same risk of motivated responding is present, but in their case interviewing is likely to lead to more passive responses. The examiner focus group used three examiners who had worked together for many years: as it was their shared experience that was of interest, and there was no hierarchical structure present, this was not considered problematic.

Procedure

For all the focus groups, the aims of the research were clearly shared with participants, but the scene-setting was deliberately vague in order to elicit broad areas of concern rather than points of specific detail. For the pupil focus groups, two researchers were present, one researcher acting as a facilitator and the other as an observer. For the teacher and examiner focus groups, an expert on the technical aspects of the topic was also available, where the participants sought clarification. The expert mode was deliberately limited in order to allow the participants to explore their own conceptions of more flexible examination delivery. The facilitators' role was to present unstructured questions and intervene only to maintain the

productivity of the groups. In this case, the participants would be brought back to the original topic, moved to another area that had been highlighted in the discussion, or moved on to a new topic altogether. The idea was to maintain as fluid a discussion as possible to capture key areas of concern rather than dictate them. The questions were adapted for the different groups to reflect the different roles that the teachers, pupils and examiners play in assessment, and were refined after each focus group session.

The examiners were presented with two scenarios for on-demand testing, and asked to think through and discuss the implications of each. In scenario 1, tests would contain a built-in 'anchor' of around ten items, which would be repeated across series in order to maintain standards. The candidates would not be told which items comprised the anchor and the items in the anchors would not contribute to their scores. In scenario 2, each candidate's test would contain one or more randomly allocated pre-test items from an item-bank. The candidates would not know which items these were, and they would not contribute to their scores. The pre-tested items would then constitute the live items for the subsequent session.

The pupil focus groups were initiated by the facilitator asking for ideas on what on-demand testing may mean, and what advantages and disadvantages of different levels of on-demand there could be. They were asked to focus on the frequency of testing they would like to see, the issue of re-sits, who should choose when they are ready, how different systems would affect the way in which they prepare for examinations, the involvement of parents and whether they would feel pressurised by different levels of on-demand. Following this general discussion, they were presented with some aspects of the scenarios of on-demand that had been presented to the examiners and were asked to comment on the fairness of these scenarios and whether they thought they could induce cheating.

The teacher focus groups began with a prompt on whether three test windows a year are enough and whether some students are more ready than others at these set test windows. Different levels of flexibility were then outlined and advantages and disadvantages discussed. In particular, the teachers were asked to comment on the perceived impact of changes on teaching and learning, school logistics, and the way in which pupils and their parents may react to the changes. Following this general discussion, they were presented with the same aspects of the scenarios of on-demand as the pupils and were asked to comment on the fairness of these scenarios and whether they thought they could induce cheating.

Consent was obtained from all participants in the study to record and transcribe the focus groups; participants were assured their responses would remain confidential and would be anonymised in any resultant reports. Dominant themes analysis was then used to analyse the data. Two researchers, one who had acted as the expert and one who had acted as the facilitator with the teacher groups, considered the transcripts separately and coded the data into themes. They then conferred and reached agreement on these themes. While this approach is subjective, it provides a useful descriptive overview of the data, and it is reassuring to note that there was little disagreement over key themes.

Results

Ten dominant themes were derived from analysis of the transcripts. The first seven themes relate to how the provision of a more flexible examination system would impact on the success of the pupils, namely: examination stress, level of flexibility, the gap between

teaching and testing, group examination preparation revision and post-mortems, parental involvement, re-sits, and good data. The remaining three themes - fairness, practical considerations, and security - relate to issues associated with the mode of testing necessary for a flexible system to be implemented. Each of the ten dominant themes will be discussed in turn.

Reducing examination stress

The dominant themes that emerged from the focus groups were unsurprisingly linked to how the provision of a more flexible examination system would impact on the success of pupils. Both teachers and pupils referred frequently to the tiredness and stress induced by examinations and how reducing this stress would lead to better examination results. There appeared to be three major sources of stress: the amount of teaching to be done before the examination, the compressed nature of the examination timetable and uncertainty about results. Both teachers and pupils referred to panic revision and cramming as they ran out of time on syllabuses:

> *'we rushed through three lessons in one and a mock and no extra practice, we could have done with a fortnight.'*
> Karen, Science teacher

Part of the problem was identified in the difficulty of planning teaching for modules with unequal weights. Some can be taught in plenty of time for the examination while others are inevitably rushed. Another aspect of the problem lay in dealing with unexpected absences:

> *'whilst maybe four or five out of the eight classes were ready at the beginning of November we might have had a second sitting for the classes who have missed lessons due to trips and fire bells.'*
> Karen, Science teacher

When it came to the examinations themselves, the teachers were particularly aware of the stress induced by the examination timetable on their weakest candidates, whose retakes were timetabled on the same day as examinations they were entering for the first time. Even a couple of days' grace, they felt, would make the difference as retake pupils were described as 'exhausted' and 'frazzled' by the experience. The pupils agreed with the need to space out examinations, and were unanimous in preferring a modular system, feeling that the alternative is:

> *'too stressful 'cause you've got all the other exams at the same time.'*
> Sophie, year 10

The modular system also means that some results are in the bank – which reduces the stress of the final examination period:

> *'you already know all your grades so you know that it takes the pressure off if you've done well in the other ones.'*
> Ellie, year 10

To mitigate the stress of examinations, pupils and teachers referred to the therapeutic process of discussing the examinations after they had occurred and a sense of group solidarity that develops amongst the pupils:

> *'it's still like scary to take it all together but you know like everybody's in the same boat, you're all doing it at the same time.'*
> Jane, first year undergraduate

Level of flexibility

Given the popularity of the current modular system, it was not surprising that the offer of more examination dates was welcomed. Specific gaps were identified for certain syllabuses where an examination opportunity would have made sense, and there was general enthusiasm for a system which could offer assessment windows every two months. Separate Science candidates, for example, follow a different schedule to Single Science candidates, but they take the same core modules with the same choice of examination dates. Different examination dates may allow the demands placed on these different streams to be better managed. The teachers also felt that if more dates were available the dates they chose could be fine-tuned to allow for different rates of progress by different classes.

There was little appetite amongst the pupils, and outright opposition from the teachers, however, to any system that would allow pupils within the same class to take examinations at different times. Most pupils, particularly those with the most maturity (the university students), saw the potential pitfalls of being offered this utmost level of flexibility and choice. Comments did range, however, from the optimistic,

> *'you'd know when you're ready and you'd probably pass,'*
> Alice, year 10

to the foreboding,

> *'you will get the people who will do it, you know for the short term just go "oh, can't be bothered" but then really suffer for it in the long term when they get to their exams.'*
> Sophie, year 10

The pupils were worried about the level of responsibility implied by too much flexibility:

> *'I think having to choose a date is just more stress than there needs to be, a fixed date you've got something to aim for,'*
> Sam, year 10

and agreed that having a fixed date to aim for reduced their stress levels. They were also worried about whether they may get left behind:

> *'when other people go ahead I'd be left behind … … I'd feel like worse they're going on and doing exams and I'm not ready for it yet. So it's better to do it all at once,'*
> Jane, first year undergraduate

and the impact of other pupils being left behind:

> *'you're going to end up with kind of people at different stages which is*
> *going to make the teaching less efficient.'*
> Nick, first year undergraduate

Only one teacher was confident her pupils would be self-motivated enough to set their own targets; the other teachers felt that the pace of lessons would suffer as pupils would not be motivated to take their examinations early.

Teachers were more outspoken on the negative impact of a system which offered flexibility at an individual pupil level. They echoed the need for dates as targets, as the dates promote group solidarity and allow pace and focus to be maintained in achieving those targets:

> *'I think it would be an absolute nightmare I really do. What I'm looking*
> *at is if I'm looking at a class that I take, and half of them have had an*
> *exam and the other half haven't, what do I do with the half that have?*
> *'Cause they now switch off and then. I think it would have to be a*
> *whole class at a time … … I know where you're going with the*
> *flexibility but I don't think it's going to work,'*
> Brian, Head of Science

> *'you still have to teach a class as a whole. You can't let some students*
> *skip straight ahead, take a bunch of exams and move onto other stuff*
> *whilst others haven't even taken the first set of exams yet, so then*
> *you'd get some students whizzing ahead of others but there is still 24*
> *kids in the same classroom. You'd have problems with the pace of the*
> *group as a whole. So if all students in the class are heading for a*
> *common date, and we all move on together.'*
> Karen, Science teacher

All agreed that any system which involved pupils within a class taking examinations at different times was a recipe for mayhem which would achieve little or no benefit.

Testing when fresh

Taking a broader view of success as effective learning rather than good examination results, a potentially harmful side-effect of more flexible delivery of examinations is the loss of synopticity – pupils learning and being tested on discrete units may not draw on and benefit from earlier learning. This could lead to superficial coverage of topics and more emphasis on teaching to a test. The benefits of further learning having the potential to improve performance on earlier topics was only raised as an issue by the university students:

> *'once I had moved further on then, the more experience with it made*
> *me understand the other stuff much clearer,'*
> Nick, first year undergraduate

> *'I took most of my AS level modules again in A2 cause I didn't do very*
> *well at AS cause – I don't know, but it was so much easier to*

> *understand once you'd done the whole topic, so it was pointless me*
> *taking the modules in the first place, you know.'*
> Jane, first year undergraduate

but even amongst themselves they disagreed:

> *'erm well I quite liked it when I could actually move on from something*
> *that I'd understood and I didn't like to keep going over things when I*
> *was getting it so I would have kind of liked to have been able to move*
> *on to the next stuff*
>
> *Yeah, just do an exam and move on'*
> Tom and Emma, first year undergraduates

The GCSE students certainly wanted to be tested while their learning was 'fresh':

> *'it's much easier to do it when you've just done a topic so then you can*
> *just quite easily do the test and you won't have to revise as much*
> *'cause you'll be able to remember everything you've been taught,'*
> Claire, year 10
>
> *'it's nice and easy, I've got, you know, my modules and so I only have*
> *to revise a small bit and then there's the test.'*
> Luke, year 10

One pupil even described the process of moving on and then taking an examination on the earlier content as confusing. Synopticity was not mentioned by the teachers, although such concerns are probably related to the level of study, the subject, and possibly even the syllabus. One teacher at least, however, was unconcerned about the gap between any teaching and testing, feeling that effective revision could compensate.

Group examination preparation, revision and post-mortems

A key aspect of success in examinations was identified as effective revision. Neither pupils nor teachers wished to sacrifice class examination preparation, revision and examination post-mortems in order to achieve flexibility. While most would prefer to revise alone, all saw the need for some group revision:

> *'it would be nice to know that you could revise with a group of other*
> *people if you get stuck.'*
> Jane, first year undergraduate

If pupils were entered at different times for an examination there was a worry that this group revision would suffer:

> *'they'd give revision sessions but I think if only one of you's doing it*
> *then I don't think a teacher would stay behind just for one person,'*
> Alice, year 10

and the lessons would suffer:

'and then other people are doing revision and other people have still got to learn.'
Alice, year 10

If everyone was given a different examination paper the pupils and teachers worried that the examination post-mortem would suffer:

'you can't actually go, you can't give meaningful feedback to them if every question paper in the room is different,'
Brian, Head of Science

but it was unclear why this post-mortem was felt to be important. Some of the pupils assumed, for example, that they would receive different mock examinations – this is unlikely to be the case.

Parental involvement

One teacher felt that one aspect of the success of their pupils was down to effective communication with parents, a process that would be made more difficult with flexible assessment:

'in this school parents are pretty much on the kids' cases so if they know the exam dates they can be sure they're revising at home, and I think that's crucial too, and I think it has a big impact on our success.'
Veronica, Key Stage Co-ordinator

From the pupils' perspective the level of involvement and understanding of their parents varied:

'mine don't really understand the situation at the moment let alone with them happening throughout the year so they don't really know when we're doing the actual thing from when you're doing the mocks and when we are doing practice ones,'
Claire, year 10

'…mine, mine are just like yeah ok, revise and then they don't really get it, 'cause they don't have a clue what I'm learning about at any point and they don't have a clue when I have tests

'Ah no, no, no, my mum, my mum gets involved.'
Sophie and Luke, year 10

Accountability to parents, however, was real to both pupils and teachers, and this aspect did raise concerns in relation to a more flexible system:

'erm, I'm not sure, I think she'd probably assume that I'd just put it off until the very end of the year and then have a panic revision and try and cram,'
Luke, year 10

> *'the people who do the same option would do the exam at the same time. Otherwise we would have parents on the phone, a nightmare.'*
> Brian, Head of Science

Re-sits

More flexible examinations offer the opportunity for more re-sits. In general, the pupils didn't feel this would change their attitude towards re-sits in any way:

> *'it's more like a safety net. It's a good thing to have, just so you know that if there is some sort of mitigating circumstances then you can always do it again some other time rather than just completely fail,'*
> Nick, first year undergraduate

> *'if you haven't passed it in like three times then you just know you're never gonna pass it,'*
> Claire, year 10

One pupil, however, felt that more re-sit opportunities would have an adverse effect:

> *'well if you had unlimited you'd just say, always think, yeah but you'd always think, you'd never try as hard as you can in any of the tests 'cause you'd always think if I screw it up this time I'm gonna just do it next time.'*
> Matt, year 10

The quality of feedback

The final aspect of success, related to a more flexible examination timetable, was the availability of good results data. One worried that too much flexibility would lead to poorer data. If different classes are taking examinations at different times:

> *'heads of departments in Key stages will find it really difficult to have an overview of the year group.'*
> Veronica, Key Stage Co-ordinator

Different tests for different pupils could also impact on the perceived quality of data:

> *'is it fair to re-organise teaching groups when they didn't all take the same exam?'*
> Karen, Science teacher

Others, however, were excited by the prospect of more data:

> *'I'm just thinking actually it's a great way to show progress…,'*
> Neil, Deputy Head

especially if the data is timely:

- 10 -

> *'I would like to see that. 'Cause I think there's quite a delay from when*
> *they actually do the exam and they get their results.'*
> Brian, Head of Science

Overall, it would seem that more flexibility would be welcomed, certainly at a year group and stream level, and, with more reservations, at a class level. This flexibility would allow the teachers to plan more effectively and reduce the stress placed on their pupils. More flexibility, however, cannot be achieved without certain changes being made to the way in which examinations are currently delivered. To what extent are teachers and their pupils prepared to accommodate these changes?

Fairness

A very flexible system might involve different pupils taking different examinations in the same subject at the same time. The pupils were slightly unsettled by this, but didn't appear to hold strong views:

> *'if at the end of it, say, me and my friend did a different exam but she*
> *got like much higher marks than what I did I'd probably be like 'oh I had*
> *a harder exam', probably try and blame it on the exam,'*
> Ilona, year 10

> *'the grade boundaries always change anyway so it's not a huge deal,'*
> Sophie, year 10

> *'it's the same at A-levels when you took another module, and when you*
> *re-took. There are different questions there so no really.'*
> Grace, first year undergraduate

Different papers for different classes aroused more suspicion amongst the teachers, however:

> *'and it would be seen as "oh you got the easy paper that's why you got*
> *better marks". It would give us another thing for parents to hit us over*
> *the head with.'*
> Brian, Head of Science

The examiners welcomed the idea of different questions being used on different papers in the same session so some questions could be pre-tested, even if those questions did not contribute to the pupils' final marks:

> *'if you ask me for my ideal world I'd construct tests entirely from pre-*
> *tested items,'*
> Kevin, examiner

> *'if it's only one item per test and if it's the only way that we're gonna get*
> *pre-testing then I'd be happy.'*
> Kevin, examiner

They did note, however, that testing time placed limits on the amount of pre-testing that could be achieved in this way. Teacher views on pre-testing in live tests were mixed:

'it concerns me a little bit in the fact that, yeah you tell the children you're looking at a mark a minute, suppose they get stuck on one of the questions that gains no marks. And that will then have an adverse, a disproportionate adverse effect on their overall grade,'
Neil, Deputy Head

'I have to say conversely I wouldn't have a problem with that at all. If you were really talking about one or two questions very short ones that wouldn't get marked I wouldn't have a problem with that at all.'
Judith, Science teacher

One teacher felt that the kids would have to be kept in the dark about this, however:

'I wouldn't have a problem with that no. We helped you with the pre-testing last year. No we would be happy with that, as long as the kids weren't aware that it wasn't being marked, we'd keep that very quiet.'
Brian, Head of Science

It seems very unlikely that keeping it quiet would be more widely acceptable.

Security

More re-use of questions may be required in order to achieve more flexibility in the examination timetable. Examiners were sanguine about this:

'so they've got to learn the science to answer the questions – even if they've seen them before…'
Jacob, examiner

while the pupils professed to only remember questions that they couldn't answer correctly,

'the only ones that you do remember are the ones that you think you got wrong or that was really hard.'
Luke, year 10

Some felt that more re-use of questions would not be so different from the present situation,

'…I mean, things like RS, you've got every single question is almost identical every year, and you've got past papers and past papers and our teachers just says you know, 'this is a question that always comes up – learn it', so, I think people would say 'what did you have' 'what was the answer' all that but, if they don't actually get back the answers within the week then it's not really too much of a problem 'cause they might be telling them the wrong answer.'
Sophie, year 10

The examiners felt that trust in the teaching profession was key:

> *'well in the same way that when we had pre-tests within schools they were administered by someone from the board the teachers from the school were sort of allowed in the room they could cast an eye over a pre-test to see what direction things were moving,'*
> Kevin, examiner

> *'I think so ninety-nine percent of the teachers are professional… and it doesn't matter what system you operate if a teacher's determined to be a rogue… …he or she will be a rogue, and I think you've got to base any system on the fact that teachers are professionals … if you take away that assumption you might as well give up… seriously.'*
> Kevin, examiner

It may be, however, that as retired teachers, the examiners' view was of a teaching profession in an age before accountability and social networking, but one examiner who was also still a practising teacher agreed:

> *'the security of an exam paper will always be compromised if somebody wants to compromise it,'*
> Neil, Deputy Head

Other practising teachers were less complacent, however:

> *'you'd be putting the teacher in an awkward situation. Where the teacher would know they could go to another colleague, another school and get some advantage for their pupils who they want to do better. So we need less temptation almost rather than more temptation,'*
> Adam, Science teacher

Technology was seen to play a role in the heightened need for security:

> *'even if papers are not photocopied and posted to everybody, ideas still transfer around the country relatively easily…'*
> Karen, Science teacher

> *'… we all have networks, associates and friends in other schools,'*
> Karen, Science teacher

> *'Yeah you could just whack it on Facebook or something like that. You could get the questions pretty easily then.'*
> John, first year undergraduate

Practical Considerations

More flexible testing would require careful timetabling. Many comments related to systems that would be unworkable:

> *'"sorry miss, I'm missing your lesson I've chosen to have my test then,"'*
> Sophie, year 10

> *'ah, a minute's notice no, it wouldn't work in school, 'cause you've planned to do this and guess what they've all gone to Colchester Zoo,'*
> Neil, Deputy Head

> *'people turning up on the wrong day for exams, people think they are going to enter for an exam when they are not,'*
> Neil, Deputy Head

> *'if we were taking out half a year group, only the triple award scientists and that would then leave half classes in French, Maths, English, we would have absolute mayhem.'*
> Brian, Head of Science

Teachers felt that the only way that the system could work was with examinations being taken in lessons, given good inter-subject co-ordination and careful forward planning:

> *'you could get round that by being able to take the exams in the classroom so you wouldn't move them all,'*
> Karen, Science teacher

> *'schools work better with dates in diaries,'*
> Neil, Deputy Head

> *'flexibility is great, but it needs to be planned flexibility,'*
> Neil, Deputy Head

> *'we'd have to know well in advance if they were going to be able to choose a date,'*
> Judith, Science teacher

> *'I would say that you wouldn't be able to schedule science exams independently of other departments.'*
> Judith, Science teacher

While these suggestions seem to offer an attractive solution, they do depend - assuming that more flexible assessment can only be delivered on-screen - on computers being available in classrooms.

Discussion

The findings reported here can be treated as fairly robust for the delivery of a more flexible examination structure for GCSE Science, for schools in the state sector. The sample size is small, but the level of redundancy in the data suggests that it is adequate. More caution must be exercised with any generalisations to other sectors of education, subjects or levels. Overall, the consensus is that more flexibility would be welcomed in planning examination sessions for year groups, streams within year groups, and even classes within streams, provided that examinations can be taken in class time, and any technology required is available in that class time. Schools could exploit this flexibility to prepare pupils more fully for

their examinations, test what is learned in a timely fashion, and reduce the stress of examinations for their pupils.

There is very limited support for any form of personalised approach which would allow pupils in the same class to progress at different rates. This is seen as a recipe for chaos which would be a nightmare to timetable and to teach. Rather than ensuring that no child is left behind it could end up isolating pupils and depriving them of the support they need. Motivation, morale and results would all suffer.

The key concern of teachers, which is not necessarily shared by their pupils or by examiners, is the need to maintain a certain level of security for the examinations. This concern appears to be heightened by the potential for mass communication that technology offers. Security leaks will spread quickly, and teachers will talk. Security concerns outweigh those of fairness. There appeared to be no serious concerns, for example, regarding the delivery of different tests to different pupils or classes, or even the pre-testing of questions in live tests, although it is preferable that these questions be marked. It appears, therefore, that the main challenge is designing a system that has enough overlap between test-takers and tests to satisfy a test-equating design, but that also maintains an acceptable level of security.

Christopher Wheadon
June 2009

References

Baird, J.-A. (2007). Alternative conceptions of comparability. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.

Gilbert, C. (2006). *2020 Vision: Report of the Teaching and Learning in 2020 Review Group*. Nottingham: Department for Education and Skills.

Newton, P. (2005). The public understanding of measurement inaccuracy. *British Educational Research Journal, 31*(4), 419-442.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus Groups: Theory and Practice*. Newbury Park, CA: Sage.

Tattersall, K. (2007). A brief history of policies, practices and issues relating to comparability. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.

Wheadon, C. B. (2008). *Maintaining standards in UK high-stakes examinations using Item Response Theory (IRT) methods of test-equating*. Guildford: Assessment and Qualifications Alliance.

Wheadon, C. B. (2009). *It's a long, long time from November to June: An investigation into increasing the flexibility of delivery of high-stakes general qualifications in England through an Item Response Theory test-equating approach*. Guildford: Assessment and Qualifications Alliance.