COMMISSION
ON SCHOOL
REFORM

Challenge Paper

Assessment and Equal Opportunities in Scottish Secondary Schools: the role of examinations and coursework

2021

Membership of the re-convened Commission on School Reform is as follows:

- **Keir Bloomer (Chair)**: Education Consultant and former Director of Education
- **John Barnett:** University Court Member, retired finance professional and former Parent Council Chairman
- **Helen Chambers:** Deputy Chief Executive of Inspiring Scotland
- **Jamie Cooke:** Head of RSA Scotland
- **Carole Ford:** Former Head teacher
- **Jim Goodall:** Former Head of Education and Community Services at Clackmannanshire Council and current Lib Dem councillor at East Dunbartonshire Council
- **Anna Hazel-Dunn:** Headteacher Royal High Primary
- **Frank Lennon:** Former Head of Dunblane High School
- **Cllr Paul McLennan:** SNP Councillor in East Lothian
- **Ross Martin:** Economic Agitator with experience in a variety of roles in education-including school teacher, voluntary tutor, chair of local education authority, member of college board & university court, Chair of a charity, and, most importantly, a parent and former student.
- **Lindsay Paterson:** Professor of education policy in the School of Social and Political Science at Edinburgh University
- **Lesley Sutherland:** Board member, the Centre for Scottish Public Policy

Please note that all members of the commission participate in an individual capacity and that the views of the commission do not represent the views of any other organisation to which the individual members belong.

Key Points

- **Scottish assessment policy has developed haphazardly without any coherent philosophy on either exams or course-work. There is therefore an urgent need for a full-scale, public review of Scottish assessment.**
- **Assessment by means of common standards has helped to guarantee the quality of Scottish secondary schooling. It has thus been at the heart of widening opportunities. For most of the past century, that assessment was by means of exams.**
- **Assessment of course-work has partly replaced exams in the past three decades. Course-work can test skills and understanding that exams miss – such as deep understanding, the ability to collate information from a variety of sources, and solving problems that require lengthy thought.**
- **The means of study for course-work has the further advantage of encouraging discussion and collaborative working.**
- **Exams have unique strengths that cannot be assessed by coursework. They test focus, memory and thinking quickly.**
- **Studying for exams can be an effective way of consolidating learning because it trains the memory.**
- **The growth of course-work has required a shift to marking by students' own teachers. This cannot be anonymous, and so risks unintentional bias against social groups who typically have low attainment.**
- **Anonymous marking of exams by markers who are external to the student's school is much less subject to bias.**
- **Both exams and assessed course-work have an important role in any system of assessment that is valid and fair.**

What's the problem?

Scottish school exams have been cancelled for the second year in a row. Though the reason is the present health emergency, the decision has provoked a wider debate about the place of exams in education. Several critics of exams have taken the opportunity to question their role. The Scottish government's International Council of Education Advisers has called exams 'essentially an out-of-date 19th and 20th century technology operating in a 21st century environment of teaching and learning'. In their critique, they claim that:

> "the capacity to apply learning creatively in unfamiliar contexts is increasingly the kind of high-value skill demanded by the workplace of the future. Traditional examinations are not capable of making such assessments on their own."

The Council was amplifying a commitment made many months previously by Education Secretary John Swinney, who said in July 2020 that:

> "there are very valid arguments for taking forward the assessment of the achievements of young people through mechanisms other than by having exams in the format that we have."

The Council was drawing on a long history of criticism of exams. Exams are said to lack validity in the sense that they do not test learning that truly matters. It is alleged that they cannot assess the capacity to think creatively, to explore ideas in an open-ended way, to find out things independently, or to engage in debate. It is also claimed that exams require the wrong kind of motivation. Instead of encouraging students to develop an intrinsic interest in a subject, exams are claimed to focus students only on getting a certificate.

There is a tendency in this debate to mix up two different matters. One is the question of exams as against what is usually called course-work, for example essays or projects. The other is about who marks whatever assessment is done, whether externally to the school or not. In practice, the growth of course-work assessment in recent years has depended on its being marked by students' own teachers, in other words internally, a change that has very significantly increased the workload of teachers. But the two questions are in principle separate. So we start by considering the importance of external assessment before returning to the question of exams versus course-work, and related questions about how to mark students' work in unbiased ways.

External assessment in Scottish secondary education: a guarantee of common standards

Against all the criticisms, the history of Scottish secondary schooling illustrates how exams can widen opportunity. Scotland adopted a system of national assessment quite early by European standards. In 1888, the school inspectors persuaded the government that externally set and externally marked assessment was needed to ensure the quality of the secondary schools that were emerging for the first time. The development of a full secondary system over the following four decades then depended crucially on the standards set by this assessment. The number of candidates rose rapidly, reaching a temporary peak of about 20% of each age group in 1924. Economic retrenchment then led to a severe restriction, to only about 5%, but the growth resumed after the late-1920s, reaching 10% in the early 1950s, 25% the 1960s, 66% in the 1970s, and more than 97% in the late-1990s and after.

At the core of the exams was curricular breadth. From 1908 until 1950, to gain the School Leaving Certificate, a student had to pass a specified group of subjects. Although the regulations varied, the group generally had to include English, mathematics, a language and a science. The legacy of this curricular structure lasted well into the second half of the century. Thus Scotland's reputation for the breadth of its curriculum depended on its system of exams.

Throughout, the original purpose of defining the meaning of secondary education remained – that is, ensuring that secondary education was at a higher standard than primary education, that it gave people their first experience of specialist study, and also that it prepared them adequately for life beyond school. Achieving this purpose required successive reforms, as proper secondary courses were extended to everyone following the introduction of comprehensive schooling in the 1960s. First Ordinary Grades after 1962, then Standard Grades from when they were phased into schools between 1984 and 1991, the Intermediate courses after 1999, and finally National 4 and National 5: all of these have been used to indicate the completion of a course at the level expected of mid-secondary school. The constant throughout has been the Higher Grade, as the main route into post-school education and the professions. Although there have often been disputes about whether the standard of a Higher today is the same as it used to be, the very stability of that level has been a symbol of Scotland's general preference for a coherent, national standard in its secondary courses.

That is then the first role which external assessment of this kind has played. It has helped to define Scottish education's integrity. As a result, it has been the means by which opportunities have slowly been widened. The first instance of this role in ensuring fairness was in relation to sex. When the system was inaugurated, Scotland was still four years away from even properly admitting women to university. But the new assessments and resulting certificates rapidly enabled young women to demonstrate their achievements as never before. By the eve of the First World War, a majority (57%) of the candidates for the leaving certificate was female. This never fell below 40% in the next half century, and equalled the male proportion from the 1970s onwards. These assessments thus became the means by which female students' potential could begin to be fairly measured for the first time, well ahead of the legislation in the 1970s that sought to guarantee equal opportunities.

The same was true for Catholics, Scotland's largest minority following immigration from Ireland and elsewhere. Through legislation in the aftermath of the First World War, the state took over the responsibility for almost all Catholic schools, with the effect that Catholic children came to have the same access as children in other schools to standardised external assessment. The role of the assessments in ensuring the quality of secondary courses was thus at the very heart of the system of Catholic schools, enabling opportunities gradually to be equalised. In 1911, the proportion of leaving-certificate candidates who were in Catholic schools was only 5%, even though Catholics were 11% of the population. By

1935, the corresponding proportions were nearly at parity, 11% and 13%. The decades following the introduction of comprehensive schooling allowed attainment by children from Catholic schools to equal that in non-denominational schools. For example, in 1998, the proportion passing at least one Standard Grade was 85% in non-denominational schools and 84% in Catholic schools.

Although socio-economic inequalities in attainment and progress have not vanished in the way that sex and religious differences have, opportunities have become more equal as a result of the extension of access to standardised exams which was opened up by successive waves of reform to the structure of schooling. As with sex and religion, that process started in the first few decades of the century, but became most visible after the coming of comprehensive schooling. Between the mid-1970s, when comprehensive schooling was being inaugurated, and the late-1990s, the proportion achieving one or more Standard Grades in the least advantaged socio-economic groups slowly caught up with the already high attainment of the most advantaged. For female students, this change was from 32% to 76% in the least advantaged, while the most advantaged had a stable percentage in the high 90s. For males, the increase for the least advantaged was from 26% to 69%, also in comparison to proportions in the high 90s. The gap remained larger at Higher level, but it also narrowed. For females, the proportion of the least advantaged who passed at least one Higher rose from 7% to 29%, while the most advantaged moved from 87% to 95%. For males, the corresponding percentages were 5% to 24% and 87% to 91%.

This gradual narrowing of the gap suggests that externally assessed tests are not intrinsically biased against students who live in socially disadvantaged circumstances. When educational conditions were made more equal – such as with the coming and development of comprehensive secondary schools – inequality gradually reduced. Indeed, probably the best place to deal with the persisting educational effects of disadvantage is in the early years of schooling, well before the stage of exams.

The reason this history matters is that it shows how external assessment has been used to establish the very meaning of secondary schooling as the stage when specialist study begins, guaranteeing that it is provided to a common standard throughout the country, for all types of school and all social groups. External assessment has been part of the gradual overcoming of invidious discrimination – part of the process by which a student is judged by what they can do rather than by who they are.

The balance of exams and course-work

For most of the period since 1888, external assessment meant predominantly exams. There has always been some recognition that not all important skills could be assessed by exams, notable early examples being in art and in laboratory work in science. But the large growth in the use of course-work did not start till the 1980s. So the second major issue in the present debate is between different ways of assessing students. A valuable summary of the evidence was provided as part of the UK government's review in 2013 of assessment in school exams in England, Wales and Northern Ireland. Because that context is more similar to the Scottish context than assessment systems anywhere else, these conclusions are relevant to the current Scottish debate.

There are two key ideas in the evaluation of any mode of assessment – reliability and validity. Reliability refers to whether a candidate would get the same result if the assessment was repeated. Validity is whether the assessment tests the objectives of the course. It is generally agreed that the strongest reliability is obtained by the most tightly standardised tests, such as the tests of cognitive ability that are used in research on intelligence. The problem is then that tightly standardised tests tend not to be valid in an educationally useful way.

For example, we can design a highly reliable test of students' capacity to solve a quadratic equation in algebra. There is only one correct answer, and so all mathematically competent markers will agree whether a particular student has got it right. Facts of this kind are important, and knowledge of them can readily be assessed by exams of a fairly traditional kind. Understanding mathematics needs a knowledge of multiplication, algebraic rules, and the laws of geometry. Understanding science needs knowledge of the laws of physics, the basis of genetics, or the periodic table. Understanding history needs a knowledge of dates and of the basic story of what happened. Understanding a language needs knowledge of grammar, spelling and intonation. Being able to play a musical instrument requires practical knowledge of scales, rhythm and harmony.

All these kinds of knowledge can be reliably and validly assessed by an exam. Indeed, most of them can now be easily assessed electronically online, and marked completely reliably in the same way. Facts of this kind are the necessary basis of any higher-order types of understanding, but judging whether a student knows them is no more than the beginning of a fully valid assessment.

As soon as we seek validity, we find ourselves drawn into the case against an exclusive reliance on exams. Judging whether a student has understood how to apply a specific bit of algebra either to a higher-level field of mathematics, or to a real-world problem in physics or economics, is not straightforward. The time available in a typical exam does not allow the complexities of genuine understanding here to be explored. It might also be said that if we want to assess whether a student can apply such ideas in the real world, then the artificiality of the time-pressured exam might not be the best way to do it.

Out of these concerns came the growth of course-work assessment. Originally, that was confined to subjects where an inescapable practical core had to be included – for example, art, music, and woodwork. But it then was extended more thoroughly than it had been previously to laboratory work and fieldwork in science, and to similar tasks in social subjects and literature. The table on page 10 shows the proportion of marks that are awarded by exam at National 5 and Higher in the twenty most popular subjects. The only one of these subjects that has no course-work assessment is mathematics. Only physical education at National 5 has no exam. For most subjects, exams still contribute the majority of marks, the exceptions being subjects with strong practical components – physical education, art and design, music, drama, woodwork, and cookery. The only other of these main subjects with a majority of marks from course-work is administration and information technology, where there is a large practical component from computing.

The case for this kind of extension of course-work assessment is strong. Worthwhile understanding of any subject in the curriculum requires not only a sound basis in facts, but also the higher-order understanding that puts the facts together. This can best be assessed by projects, experiments, and essays. However, none of these arguments in favour of course-work shows that exams are redundant, because well-designed exams not only test other important aspects of knowledge and skill, but also can address several of the features that make course-work useful. Both modes have strengths. In an extensive review of research on course-work and exams in higher education, Richardson (2015) notes the advantages of each. All these features would be as relevant to the senior years of secondary school as to universities.

- Course-work encourages students to develop a deep understanding, to learn how to collate information from a variety of sources, and to solve problems that require lengthy thought. Course-work is probably also better than exams at encouraging motivation over a long period. The work involved in preparing an essay or a project can involve discussion and collaboration, both valuable features of consolidating learning as well as useful skills for life.

- However, revising for exams, too, can be motivating, and indeed a strong argument

for exams is not the assessment itself but the kind of study that is required to prepare for them, especially the training of memory. Memorising poems leaves a legacy that can last long after the exam is over; committing algebraic formulae to memory trains the mathematical mind; being able to summarise succinctly the complex reasons for the Highland Clearances discourages simplistic political rhetoric. In any case, training the memory is a key skill, despite claims that we do not need to remember anything in the world of the internet: one reason is that, to make sense of anything on the internet, we need memory to fit new facts into a framework of understanding. Exams also test the capacity to think quickly and to focus, both of which are as important skills for life as any of the skills tested by essays and projects.

- The validity of exams has been questioned on the grounds that they can narrow the curriculum. This is often expressed as a fear that teachers 'teach to the test', instead of covering the range of understanding that education ought to encourage. One reply to this is that course-work is not free of such distortion, because course-work, too, can become formulaic. But the main response is to insist that exams be designed so that teaching to them encourages deep learning. The exam must sample widely from the course content, to ensure the whole course is taught, the questions should not be so formulaic that it is possible to drill pupils in how to respond, and questions should not require prepared answers, such as in the writing elements of English, history and foreign languages. Exams ought to be sufficiently long to provide a full opportunity for the pupil to display their understanding and knowledge.

- A different kind of question about validity relates to whether the work is genuinely the student's own, which has sometimes been described as the question of authenticity. Whereas there are educationally legitimate reasons to give credit for skills that cannot be demonstrated in an exam, there are no grounds for rewarding plagiarism, or for attributing to a student's own work the opportunity to discuss a project with a parent who knows a lot about the topic. Outright plagiarism can be dealt with to some extent, for example by electronic means of detecting it, but there is no easy way of avoiding the more complex kinds of inauthenticity. Even if all course-work was completed in school, with no copies taken home, that would still not prevent a student from discussing the work with people outside the classroom.

Richardson also summarises the evidence relating to a familiar claim – that female students prefer course-work whereas males prefer exams. Most of these claims are speculative rather than based on measurable evidence. For example, people have noticed that the growing use of course-work in school assessment has coincided with the emergence of an average attainment gap in favour of female over male students, and have suggested that the two might be linked. In fact, though, the explanation is probably in study habits rather than in the modes of assessment. Any student who studies hard over a long period of time will tend to do better in both exams and course-work. As it happens, a greater proportion of female than of male students study in that way. The resulting difference in attainment is then more appropriately attributed to study habits than to the modes of assessment as such.

Marking

Moreover, although course-work may solve some problem of validity, it exacerbates the difficulty of reliability, and also potentially introduces a new version of invalidity, the bias of markers. Fortunately, the ways of dealing with both of these are similar.

Reliability in marking essay-type questions is a notorious problem, but three procedures can help to mitigate it. One is to have several people mark independently, and then hold a discussion among them to resolve differences. Essentially this is what is done for marking research dissertations in UK universities, but it would be wholly unaffordable to be used routinely for marking course-work at any level of education. A version of this is what has been called a 'community of practice', as might gradually develop from a group of teachers who have worked together for years, and who have well-developed ways of inducting new teachers into the group. The problem is that, although the community might agree with each other, it might drift away from practice in other communities. So the third principle has to be clear marking schemes, careful training of markers in them, and re-marking of a sample of scripts anonymously by external inspectors. All of this is standard practice by the Scottish Qualifications Authority, and has been since the beginning of school exams.

It turns out that training in an agreed rubric is also an effective way of dealing with bias. Bias does not arise where the marking is anonymous, but the sheer scale of course-work assessment now has required that most of it is marked by students' own teachers. It is then almost impossible for the marking to be anonymous. Even if names are removed from scripts, hand-writing and other idiosyncrasies are recognisable to a teacher who has known the student for many years. Few subject departments in schools would be large enough for marking always to be done by teachers who do not know the students individually. There is no scope in most assessment for anything analogous to what is done to avoid bias in auditions for orchestras, which have to be conducted behind screens so that the judging panel has no idea of the sex or ethnic background of the applicant.

There is strong evidence of bias in marking that is not anonymous, which takes two forms, and which does not have to be intentional to be a problem. One is bias against or in favour of social groups that, on average, tend to have untypical attainment. These patterns vary over time, and so the bias is itself likely to vary, but there is good evidence of bias against students in socially deprived circumstances. There is also bias in favour of certain groups. Stereotypical views such as that Chinese pupils are good at maths, girls are better at foreign languages, or boys are more interested in physics, exist amongst teachers as well as in the general population.

The other main kind of bias in non-anonymous marking is what has been called the 'halo effect', in other words expectations based on the individual candidate's past performance. Students who did well in the past are expected to do well again. The halo effect may interact with other kinds of bias. For example, teachers' beliefs about the mathematical aptitude of boys and girls has been found to be one factor in explaining higher mathematical achievement by boys.

There are further problems. Appeals against grades are much more difficult to arrange when a large component of the assessment is by teachers' marks. To enable students to appeal against these, there has to be a mechanism by which external assessors re-mark the work, because in most schools there is no-one sufficiently independent of the original marking process. But having to invoke external assessors in this way means that students who appeal are assessed by potentially different standards to those who did not appeal. If a teacher grade from course-work is very different from the grade based on the anonymous marking of an exam, then the unavoidable risk of teacher bias would have to lead to giving precedence to the grade derived from the exam, on the grounds that, in such circumstances, the risk of distortion from the exam grade is far lower than the risk from the teacher grade. However, if the teacher and exam grades are quite close, then there would be a stronger case for an appeal, because that would suggest that the teacher had successfully guarded against bias. Either way, however, this series of decisions would have the incidental effect of placing greater emphasis on the exam than on the teacher assessment, defeating the purpose of the dual approach.

There is also the problem that teachers can be put under pressure to award high grades, a risk that

has been shown to be acute in the USA where teacher grading plays a much greater role than in European systems. Anonymous marking is thus a good defence of teachers' professionalism.

Theoretical critiques of exams – such as from the Scottish government's International Council of Education Advisers –tend to ignore these difficult practical challenges. The sheer growth of course-work threatens the integrity of school assessment because marking so much course-work externally and anonymously is simply not feasible. Externally marked exams are the most practicable way of making sure that an A in one school means the same as an A in another. This ensures that there can never be the disputes about the comparability of school results that there are in debates about standards across universities, where (except for research dissertations) the marking is almost wholly internal. Basing assessment predominantly on course-work that is marked by teachers would be unavoidably unfair.

Conclusions: the need for a full public review

External assessment has been at the heart of universal secondary schooling. It is the norm in school systems across the world, including in countries, such as Finland, which generally place less emphasis on formal assessment than Scotland.

Scotland's systems of assessment have grown haphazardly and without any long-term planning, and so the debate provided by the present health emergency is valuable. There is now an opportunity to develop a consensus around a clear explanation of fundamental philosophy. The confusion arising from the absence of such coherence has allowed the important role for exams to be put in doubt. The body responsible for assessment – the Scottish Qualifications Authority – has never set out its principles, and never explained what the unique contributions of exams and course-work are. Nor has it explained that course-work, as it has evolved, has serious weaknesses, most notably in the workload it imposes on teachers and the unavoidable risk of unintentional bias that relying on teacher marking poses.

Matters of such fundamental importance should not be decided in the heat of political debate during a health crisis, and nor should they be a matter only for the secret deliberations of councils of educational experts. There needs to be a full-scale, public review of Scottish assessment. The remit of the review should cover:

- the purpose of assessment both at the end of schooling, and in the intermediate years;
- the role which external assessment plays in ensuring equality of opportunity;
- the strengths of exams, and how exams might be redesigned to take full advantage of their potential;
- the strength of course-work assessment, and how existing course-work assessment might be redesigned to take advantage of its potential;
- what if any role there should be for external, anonymous marking of course-work;
- the different roles of different forms of assessment in different subjects of the curriculum in the senior school years, and how these differences might be reconciled with a common philosophy of assessment;
- whether and how the development of educational technology might change the nature of assessment, especially in the light of the experience of the cancellation of exams in 2020 and 2021.

Percentage of marks from exam in the most popular subjects at National 5 and Higher

Subject	National 5	Higher
English	70	70
Mathematics	100	100
Biology	80	80
Physical Education	0	50
Chemistry	80	80
History	80	73
Physics	80	80
Modern Studies	80	73
Geography	80	73
Art and Design	20	23
French	63	63
Business Management	75	75
Music	35	35
Computing Science	69	69
Graphic Communication	67	64
Practical Woodworking	30	No Higher
Spanish	63	63
Hospitality: Practical Cookery	25	No Higher
Administration and IT	42	42
Drama	40	40
Human biology	No National 5	80
Religious, Moral and Philosophical Studies	80	73
Psychology	70	67

The subjects are listed by descending order of the number of candidates in 2019 at National 5, covering the 20 most popular subjects at each level. There is no Higher in practical woodworking or practical cookery, and drama is not in the top 20 for Higher (being 21st). The three subjects at the foot of the table complete the list of 20 most popular subjects at Higher.

Sources: Course Specification documents from SQA website (https://www.sqa.org.uk/sqa/70972.html), accessed February 2021. Information on the number of candidates in each subject at each level is from Attainment Statistics (December) 2019, published by the SQA.