Coimisiún na Scrúduithe Stáit
State Examinations Commission

Report on the Trialling of the

Assessment of Practical Work

in

Leaving Certificate

Biology, Chemistry and Physics

June 2018

November 2018: this version incorporates the correction of some typographical errors noted after submission of the report.

Contents

Acknowledgments

The State Examinations Commission would like to thank the students, teachers, school principals, and other staff in the schools where the trial took place for their co-operation and the enthusiasm with which they engaged with the trial, and for their time and care in providing comprehensive and thoughtful feedback.

We also thank the examiners who volunteered to be involved with the project and their school principals and management authorities for agreeing to release them from normal duties to undertake this important work.

The level of disruption for all concerned was exacerbated by the effects of Hurricane Ophelia, which coincided with some of the planned dates for the trial in schools and necessitated rearrangement at short notice. We are grateful for the understanding, flexibility, and resourcefulness of all involved, which allowed the trial to be completed despite this difficulty.

Summary

Draft new subject specifications (previously known as syllabuses) for Leaving Certificate Biology, Chemistry and Physics have been prepared by the National Council for Curriculum and Assessment (NCCA). These specifications include a proposal that 30% of the marks in each of these subjects be awarded for a 90-minute, laboratory based, externally assessed, practical examination, with the remaining 70% being for a written examination.

The proposed arrangements were trialled in thirty schools in October 2017. Feedback was systematically gathered from all involved (students, teachers, school management, examiners, task setters, and chief examiners), and detailed quantitative and qualitative analyses were carried out on the feedback, the tasks, the outcomes, and the process. A smaller proof-of-concept trial on the possible use of digital technology in such an assessment was also carried out in three additional schools.

While the feedback from those involved was generally very positive towards the assessment of practical skills in principle, participants expressed concern about aspects of implementing it as a component of the Leaving Certificate examination. The main concerns expressed related to the disruption to the normal life of the school, the amount of work required of teachers to support implementation (and the attendant impact on their other work) and the availability of examiners. While most of those involved agreed that science practical skills were important and should be tested, there was less agreement that that this was the right way to do it. While students strongly agreed that they would pay more attention to practical work if such an assessment were introduced, there were lower levels of agreement with a statement that such an assessment *should* be introduced, and lower levels of agreement still that the assessment as they experienced it was fair. Notwithstanding this, these statements all attracted more agreement than disagreement.

Technical analysis of the quality of the assessment did not throw up any significant concerns that could not be rectified. Quantitative analyses provided *some* evidence that the assessment measured a distinct set of skills from those measured by a written examination, but not strong evidence. The lack of strength in the evidence may have been related to the availability of only a weak alternative measure of practical skills with which it could be compared, and the analysis was certainly not such as to undermine an assertion that the assessment measured different competencies from those measured by a written test. Qualitative analyses suggested that the assessment did indeed focus on the skills that had been specified for it.

Nonetheless, it was known from the outset and came into greater relief over the course of the trial that, in an ideal world, an assessment of practical work in a modern science curriculum should, if it were possible, seek to evaluate the students' capacity to apply their knowledge and skills to unrehearsed and less familiar problems. The context of the Leaving Certificate examination as an entirely externally assessed and high-stakes examination prevents that from being achieved with an assessment of this type, so we must remain clear about what this assessment can and cannot do when assessing the benefits in comparison to the costs (in the broadest sense).

The logistics and costs of a roll-out were interrogated. Total costs of delivery of the assessment itself, excluding certain headquarter costs and costs associated with preparing materials at school level, are of the order of €2.5 million annually. Including all costs would be unlikely to bring this figure above €3 million annually. This is exclusive of any capital and recurring costs that might be required to bring laboratory facilities in schools up to the required standard, to maintain them at that level, or to provide technical support to schools. Auditing science facilities in schools and estimating such costs were beyond the scope of the trial. Leaving aside costs, there are significant concerns about the capacity to deliver a full rollout along the lines of existing models of delivery of oral and practical examinations, which are already under considerable stress. Looming large among those concerns is that of examiner supply. For this and other reasons, if a decision is made to proceed, a major review is required of how all oral and practical assessment is organised, how it integrates with the other needs of the education system, and how it is to be supported by the education community at large.

The implementation challenges identified in the report and the limitations as to what can be adequately tested in an assessment of this type need to be recognised. However, the benefits of proceeding, in terms of impact on the teaching and learning of science, should not be undervalued.

While the report is most coherently read as presented, there may be readers who are not in a position to read the full report, but who nonetheless wish to appreciate fully all of the issues that must be considered when a final decision is being made regarding a national rollout of the proposed form of assessment. Such readers are advised to read Chapter 9 in its entirety, following which they can refer back to other chapters as required.

Chapter 1 Context, purpose and scope of the trial

> 'Why practical science?' ... Experimentation gives science its identity. Science uses experiments to discover the realities of the world and this practical approach seems to be as intrinsic to young learners as it is to professional researchers. The attraction of practical science seems to lie in its appeal to the senses, its surprises and its unpredictability, as much as in its power to explain. The real world is not cut and dried, and nor is practical science. Experiments do not always go as expected, and we can learn as much from unexpected results as from expected ones. And the same is true of life.
>
> *(Holman, 2017)*

1.1 Context

In April 2013, the National Council for Curriculum and Assessment (NCCA) completed work on draft new subject specifications for Leaving Certificate Biology, Chemistry and Physics. Each specification included an externally assessed practical assessment component as a part of the Leaving Certificate examination in the subject. This is a new development, as the direct assessment of practical skills has not heretofore been a part of the Leaving Certificate examinations in these subjects.

In June 2014, following consultation with the NCCA, the State Examinations Commission (SEC) submitted a proposal for a trial of the proposed model of practical assessment to the Department of Education and Skills (DES). The purpose of the trial was to gather information on whether and how such an examination component might be made to work, including: investigating the feasibility and validity of potential kinds of assessment tasks; establishing what resources are required for such tasks to be carried out in an examination context; clarifying the role that the science teacher would need to play in supporting the arrangements; exploring the logistics of running the examination, including establishing the number of students that could be accommodated in each practical examination session and the number of sessions that could be held each day, and hence the number of examiners that would be required; and exploring the impact on students, teachers and schools.

In September 2016, the DES approved the proposal for the trial. The necessary implementation arrangements were put in place, and the detailed work of the trial began. The in-school main phase of the trial occurred in October 2017, and much information – both quantitative and qualitative – was gathered on all aspects of the practical assessment. This report is based on that information and experience and is intended both to assist the NCCA in finalising its advice to the Minister and his Department and to assist the Minister and the Department in making a final decision as to how to proceed.

1.2 Assessment of practical work and the new subject specifications

The new specifications state that students of Leaving Certificate Biology, Chemistry, and Physics will be expected to understand abstract concepts and to be able to transfer them to new contexts. It states that learners will also be expected to be able to apply skills and strategies of scientific inquiry to solve problems and in participating in practical activities, they will be required to manipulate and use tools, equipment and materials safely. Students should

- develop skills in scientific inquiry, including the ability to interpret and analyse qualitative and quantitative data from different sources and to consider the validity and reliability of data in presenting and justifying conclusions;
- develop skills in laboratory procedures and techniques, including the use of technology, carried out with due regard for safety, together with the ability to assess the uses and limitations of these procedures through engagement in a wide variety of practical activities; and
- be able to discuss, evaluate and communicate the results of their experimental and investigative activities in verbal, graphical and mathematical form, using ICT where appropriate.

The draft specifications prepared by the NCCA were supported by two other documents, *Senior Cycle Sciences – Assessment Outline* and *Leaving Certificate Science Subjects – Working Towards Implementation*.

In *Working towards Implementation*, it is pointed out that the inclusion of practical assessment in the Leaving Certificate examinations for the science subjects is reflective of international developments in 21st-century science education. The thrust of such recent developments is to emphasise the importance of developing in learners the skills associated with scientific practices as well as the concepts and theories of science.

The draft specifications state that practical work in the sciences will involve learners collaborating and working in groups as they use technology to collect and analyse data, to research ideas, and to communicate their findings. The NCCA proposal for a practical examination to complement the written examination in each of the senior sciences supports a greater emphasis on student practical work in teaching and learning.

The NCCA documents acknowledged that implementing practical examinations in the sciences would be expensive and logistically challenging. The proposed timeline for implementing the new subject specifications envisaged them being introduced in September 2018[1] after a period of development of system capacity following the trial of practical assessment. It was also stated in the documentation that the draft subject specifications and assessment arrangements would be reviewed in light of the outcomes of the trial.

[1] This proposal was made at a time when it was envisaged that the first cohort studying the new junior-cycle science specification would complete junior cycle in June 2018. The corresponding target would now be September 2019. However, given the groundwork required in advance of introduction, this is no longer feasible.

1.3 The proposed model of practical assessment

The NCCA documents mentioned earlier deal with the rationale for including a practical assessment component in these subjects in the Leaving Certificate examination and explore various possible models of practical assessment, so this discussion is not repeated here. The model ultimately proposed for use in the Leaving Certificate examination and hence for trialling has the following features:

- It is a 90-minute, invigilated, laboratory-based practical examination.
- It is carried out in the school's laboratory for the subject, using the school's equipment.
- An external examiner, appointed by the State Examinations Commission, visits the school to conduct the examination; this person acts as both invigilator and examiner.
- Each practical examination session involves up to twelve candidates working independently, each assigned a different (set of) task(s).
- The tasks are based on mandatory practical activities stated in the respective subject specifications (although this does not necessarily imply that they are simply replications of such activities).
- Candidates are assessed in two ways: first, the examiner who is present on the day assesses the candidates' practical skills during the session, awarding marks 'in real time'; second, task-specific booklets that the candidates complete during the session are marked later.

The proposal is that this practical assessment component would carry a total of 30% of the overall marks for the examination, with the remaining 70% allocated to a written paper. The 30% allocated to the practical assessment would be divided evenly between marks awarded by the examiner on the day and the marks subsequently awarded for the work in the booklet. The focus of the marks awarded by the examiner on the day is on the practical skills that he or she can directly observe the candidate displaying, (specified in more detail later in this report). The focus of the marks for the booklet is on the proper recording of what was done and what was observed, on the quality of the analysis and conclusions drawn from what was observed, and on knowledge, understanding, and application of associated subject content. For the purpose of this report, the first of these two aspects (the marks awarded by the examiner on the day) is referred to as the assessment of *practical skills*, and the second (the marks for the booklet) as the assessment of *results and analysis*. Earlier in the project, these were referred to respectively as the *direct assessment* mark and the *indirect assessment* mark as working titles, but these terms do not adequately capture their form or function and so have been amended.

1.4 Scope of the trial

The purpose of the trial in broad terms was stated in 1.1 above. Having explained the proposed model of assessment more clearly, we can now be more specific about some of the questions that the trial was designed to assist in answering:

- Can a suitable range of tasks be prepared for each subject, providing an appropriate vehicle through which the target skills can be validly and fairly assessed?
- Can the candidates carry out these tasks (and complete the booklets) comfortably in the allocated time?
- Can examiners reliably assess the practical skills that they observe the candidates displaying?
- Can examiners assess twelve candidates in a single session?
- Can three 90-minute examination sessions be accommodated during a school day?
- What range of impacts is such an examination component likely to have on teaching and learning?
- What level of intrusion on the other work of the school would such an examination component have?
- To what extent would such an examination component receive the support of students, teachers, and school authorities?
- What delivery models might be feasible, given systemic constraints, including examiner supply?
- What resources (financial and other) would the SEC require in order to implement a full rollout of such a component under any feasible models identified?

It should be noted that the trial cannot fully answer all of these questions. While it can provide a comparatively complete basis to answer some of them, it can only contribute to a lesser degree to answering others. It should also be noted that, of necessity, the tasks used for the trial were based on the mandatory practical activities in the current syllabuses rather than the new subject specifications. Furthermore, the timing of the trial early in year 2 of the Leaving Certificate programme limited the extent to which relevant content had been covered by the students and therefore the range of tasks that could reasonably be used in each school. In addition to the constraints that these two factors place on the science knowledge that could be assumed, basing the tasks on the existing syllabus might also be considered to have constrained the nature of the types of skill that students could reasonably have been expected to demonstrate.

It is appropriate also to point out that there are questions that were considered from the outset to be beyond the scope of the trial, but which nonetheless need to be considered and addressed before any decision is made to roll out the proposed examination component. This includes, for example, the question of whether all schools have the laboratories and equipment necessary to facilitate the holding of such an examination component, and, if not, what level of investment is required to bring them to such a state.

An additional small-scale 'digital' trial was added to the main trial. The purpose of this was to serve as an exploration or proof-of-concept as to the feasibility of delivering the trial through a digital medium instead of the written tasks booklets. If rolled out on a large scale, such a mode of assessment could, in theory, facilitate task types and formats not possible with paper booklets, such as integrating data-logging and other electronic data-

capture tools with software for data analysis. For the most part, this report deals with the main trial only. Reporting of all matters related to the digital trial is consolidated into Chapter 7.

1.5 Making decisions about testing policy

Decisions about educational testing policy can have far-reaching consequences for students and other stakeholders in the education system. In Ireland, the Leaving Certificate examination exerts a huge influence over not only the learning experiences of second-level students, but also on their future studies, careers, and life chances. This makes it all the more important that any decision about what happens in this examination be based on a thorough evaluation of the proposition and all of its potential consequences.

Such an evaluation certainly encompasses the matter of a conservatively expressed vision of test *validity* (does the test measure what it is supposed to measure?) and *feasibility* (can we deliver it?), but goes well beyond these.

Newton and Shaw (2014) provide a thorough overview of the evolution of the concept of test validity over the last century. The book culminates with a proposed *framework for the evaluation of testing policy*. This framework provides a useful structure for those involved in formulating policy on examinations and assessment, allowing them to have comfort that they have at least considered all of the important categories of question that need to be considered when making the overall evaluative judgment as to whether to proceed with any testing programme under consideration. See Figure 1.1 for an overview of this framework

As noted in Section 1.4 above, this trial can provide partial or complete answers to *some* of the questions in *some* of the cells in this framework. Nevertheless, as this report is intended to be a key support in the subsequent decision-making process, it would be remiss not to draw attention to and explore some of the other questions. They are therefore drawn into consideration towards the end of this report.

Figure 1.1 Framework for the evaluation of testing policy (Newton & Shaw, 2014, p.186)

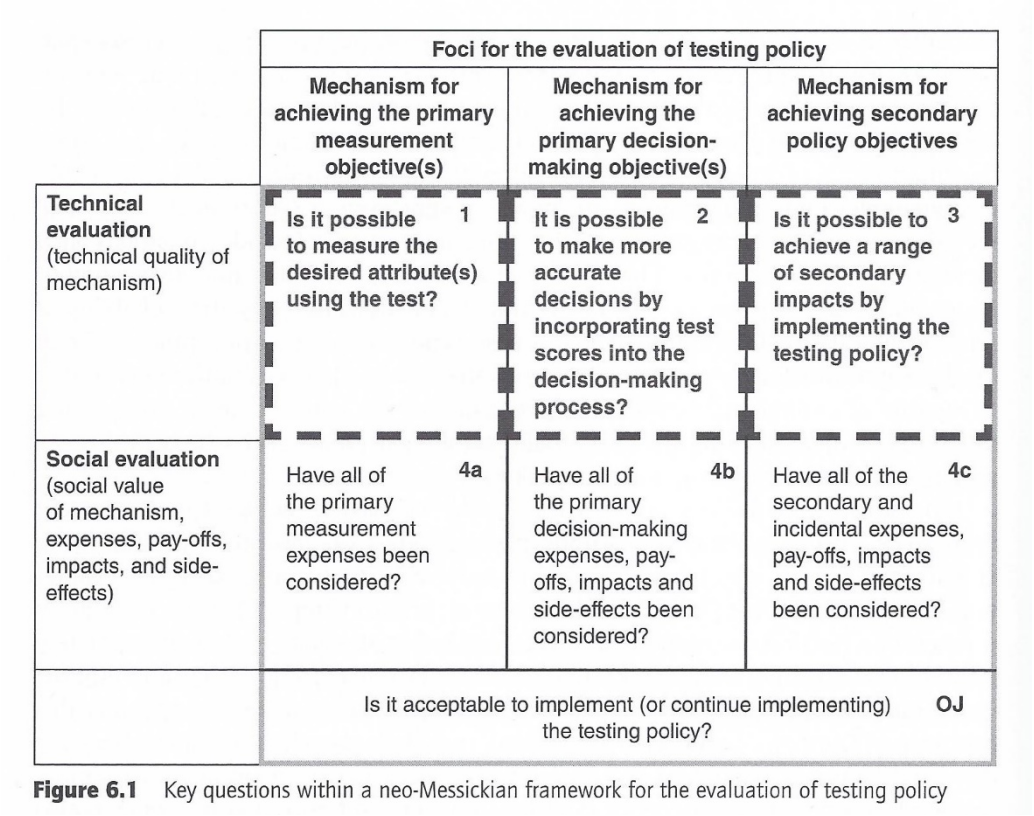


Figure 6.1 Key questions within a neo-Messickian framework for the evaluation of testing policy

1.6 Structure of this report

This report has been structured in such a way as to organise the information gleaned from the trial and deal with related issues in a way that reflects the kinds of question that must be addressed when formulating advice or making policy decisions about testing programmes, as outlined in Section 1.5 above.

Chapter 2 describes how the trialling project was planned and implemented, so that the reader can assess its soundness as a basis for any findings and conclusions drawn. The following chapters are the core of the report, laying out the information that the trial was designed to elicit in order to support a sound decision as to how to proceed – Chapter 4 deals with the information systematically gathered from the participants in the trial, offering a rich insight into how they experienced the trial and their attitudes to this form of assessment on the basis of that experience. While they may not be strictly said to be a representative sample of their respective populations, these insights can nonetheless reasonably be taken to be touchstones as to the likely views of these critical stakeholder groups in the event that those groups were exposed to a similar experience. Chapters 5 and 6 then deal with the technical quality of the assessment instrument as implemented, based respectively on the available quantitative and qualitative evidence produced by the trial.

Chapter 7 to some degree stands independent of the other chapters. It deals with all aspects of the digital trial – a parallel 'mini-trial' designed as a preliminary exploration of the potential offered by an assessment model that uses the same basic assessment model but integrates digital media with practical assessment activities in science.

Chapter 8 deals with the nuts and bolts of what it would take to deliver this assessment model on the scale required for a full national roll-out as part of the Leaving Certificate examination, insofar as this can be predicted in advance. Finally, Chapter 9 draws all of the information from the preceding chapters together in an attempt to answer the basic overall question: what does the trial tell us about whether we can and should proceed to embed the proposed assessment model into the Leaving Certificate examinations of the subjects concerned? As mentioned in Section 1.5 above, recognising that there are many relevant questions that the trial could not address, this chapter raises and briefly discusses some of these, for fear that decision-makers might lose sight of them if they were to regard this trial as the sole or dominant basis for their decisions.

Supplementary information and supporting documentation is provided in the appendices. Because they are extensive, these are, for the most part, being made available separately rather than being included in the circulation of the body of the report.

Chapter 2 Planning and implementing the trial

2.1 Project governance, oversight, management and reporting

The trial was carried out by the State Examinations Commission, in consultation with the National Council for Curriculum and Assessment, and at the request of the Department of Education and Skills. This report is a report of the Commission to the Department.

The following structures and arrangements were put in place in order to carry out the trial.

2.1.1 Trial manager

The trial was managed on a day-to-day basis by the *trial manager* who worked with the SEC's Examination and Assessment Managers (EAMs) for Biology, Chemistry and Physics in carrying out the trial. The trial manager, who is a retired DES science inspector, was appointed on a contract basis by the SEC to carry out this function.

The trial had administrative support from the SEC headquarters in Athlone and the SEC offices in Limerick.

2.1.2 Implementation group

While the trial manager managed the trial and carried out much of the work involved in implementing it, the detailed implementation of the trial was directed by the *implementation group*. The implementation group consisted of: the trial manager (chair); a co-ordinating representative of the SEC's Examinations and Assessment Managers (EAMs) for Biology, Chemistry, and Physics; the SEC's Assistant Head of Examinations and Assessment Division (AHEAD) with responsibility for the sciences; a science inspector from the Department of Education and Skills; and two nominees from the NCCA. The implementation group met every three to four weeks over the course of the trial and contributed in detail to decision making, preparation of documentation, and discussing and formulating observations and findings. The membership of the implementation group is given in Appendix A.

2.1.3 Steering group

The *steering group*, which had overall responsibility for guiding the implementation of the trial, included representatives of the SEC, the NCCA and the DES (Inspectorate, Curriculum and Assessment Policy Unit, and Planning and Building Unit). The group was chaired by the SEC's AHEAD with responsibility for the sciences. The role of the steering group was to provide guidance on broad, high-level issues and to monitor the progress of the trial in relation to its objectives and the broader implications for the system, so as to ensure that the trial provided a sound basis for decision-making in relation to a full roll-out of practical examinations in the senior sciences. It met on four occasions during the trial: near the beginning; in advance of the implementation of the in-school phase; following the in-school phase; and when a draft version of this report was available. The trial manager attended meetings of the steering group and acted as its secretary. The membership of the steering group is given in Appendix A.

2.1.4 External project reviewer

Because of the systemic importance and far-reaching consequences of any decision to roll out a practical assessment as a component of the Leaving Certificate examination in the science subjects, it was decided that the project should be reviewed by an independent external agent, in order to provide assurance to stakeholders that the project was fairly and appropriately planned and executed.

Sir John Holman, Emeritus Professor of Chemistry at the University of York, President of the Royal Society of Chemistry, was engaged as the external reviewer in September 2017, before the in-school phase of trial was implemented. The reviewer's role involved assessing the project plan, reviewing documentation to be used, observing the trial in a small number of schools, and reviewing the draft final report.

The detailed specification of the reviewer's role, along with his interim and final reports, is in Appendix B.

2.2 Preparation for the trial

There were three major aspects to the preparations for the trial: the preparation and testing of tasks; the selection and briefing of schools; and the appointment and training of examiners. The EAMs with responsibility for Biology, Chemistry and Physics prepared the tasks with the assistance of setters, as is normal practice for the preparation of SEC examination materials. The EAMs also selected and trained the examiners, having sought expressions of interest from members of existing examining teams and teachers in the trial schools. They worked closely with the trial manager through a series of regular meetings. At each stage, the trial received invaluable administrative support from the SEC offices in Athlone and Limerick.

2.2.1 The tasks

The first priority of the implementation group, to which the trial manager reported at its meetings, was the outline design and preparation of suitable tasks for the trial. The trial manager and the EAMs met separately to work on the tasks at the earliest opportunity. They had access to draft practical assessment test items that had been prepared in advance of the trial by the NCCA.

The work of planning the structure of the tasks and of writing guidelines for the work of the setters commenced following the first meeting of the trial manager with the EAMs in December 2016. Along with sample tasks prepared by the EAMs, twelve tasks were prepared for each subject for the trial – in each case by two setters working with the relevant EAM. The setters' brief required them to provide suitable draft tasks, an equipment list for each draft task, and a marking scheme for each draft task.

The preparation of the tasks was demanding for both setters and EAMs because of the timescale of the trial. Even though the SEC was able to provide some partial relief from other duties to the relevant EAMs, it proved difficult to accommodate the substantial amount of additional work required. It is clear from the experience of setting the tasks for the trial that, in the event of the adoption of practical assessment as a part of the Leaving Certificate examination, preparation of suitable tasks annually will be demanding. Some

pre-testing of the tasks will also be essential to ensure they can be feasibly executed as envisaged in the time available. It will be essential that adequate resourcing is provided for this aspect of the work.

2.2.2 The 'trying out' phase

In April 2017, when the draft tasks became available, a small-scale 'trying-out' was organised. The purpose of the trying-out was to test whether the draft tasks were suitable for the trial practical examination. Twenty-four teachers, eight for each subject, were invited to try out the tasks with a small number of their students, approximately four for each task. Teachers had to gather the equipment specified for each task in the laboratory in advance. On the day, students chose the specific equipment they needed and carried out their assigned tasks. The trying-out teachers were not requested to assess their students, nor was the marking of students' written work completed as part of the trying-out.

The trying-out phase of the trial culminated in a 'roundtable day' on 9 May 2017 when the teachers and the trial team gathered and discussed issues that had come up in the schools. The feedback provided by the teachers and their students was extremely valuable in the subsequent preparations for the main phase of the trial, as it was possible to address some of the issues identified, at least partially, before final planning commenced for the main trial in October. The following are some of the changes made to the trial procedures following the trying-out phase.

- The laboratory for the trial practical assessment should be the one designated for the subject. (Prior to this, it was an open question as to whether a laboratory for one of the three science subjects could be used to hold an examination in another one.)
- Students should be allocated a dedicated period of ten minutes at the beginning of the practical examination to read their tasks, to plan, and to collect equipment.
- A dedicated ring-fenced mark for students' clean-up at the end of an examination session was included in the marking scheme for each task.

Some minor changes were also made to the tasks.

Following the roundtable day, the tasks and equipment lists were prepared by the EAMs and brought to a final pre-translation state. They were then translated and final versions of both English and Irish versions were signed off for printing. They were printed by the question paper unit of the SEC and packed for use in the trial. Digital completion booklets were also prepared for use by any students who normally require a writing accommodation in examinations and whose normal way of working in class involves word processing.

2.2.3 Selection of schools

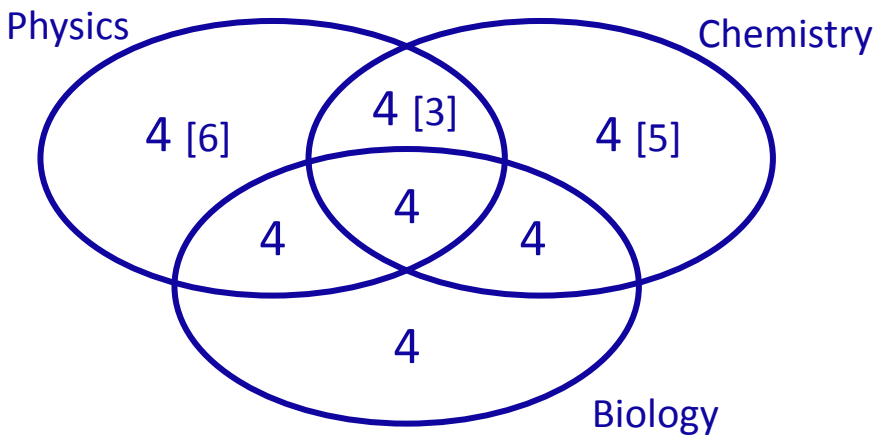
As has been normal practice for trialling projects such as this, expressions of interest were invited from all second-level schools, with the selection to be made from among those expressing such interest.

From the first invitation to participate, a high priority was set on clear and timely communication with schools. A detailed information note on the trial was included with the

invitation and schools were allowed several weeks to reply. By the closing date at the end of April 2017, more than two hundred schools had applied to participate in the trial.

The number of schools required to adequately and efficiently trial the aspects of the proposed assessment model that required evaluation was judged to be 28, to include a representative range of school types and contexts, and to include schools trialling one, two, or all three subjects simultaneously, aiming to match the schema shown in Figure 2.1. This schema results in each subject being piloted in sixteen schools, incorporates considerable experience in trialling two subjects in a school, and some experience of trialling all three subjects in a school, these latter aspects being necessary to evaluate the impact of assessing two or three of the subjects on the life of a school.

Figure 2.1: Target [& achieved] number of schools trialling each combination of subjects.



In order to mitigate the risk associated with the possibility of a school dropping out, two additional schools were selected, to bring the total to thirty. These were selected in such a way as to increase the total student participation level in Physics and Chemistry, which have lower numbers of candidates per school than Biology.

The thirty schools were selected for the trial in early May. The total number of students expected for each subject in the schools selected, on the basis of the student numbers supplied on application, was: Biology: 810; Chemistry: 432; Physics: 412; total: 1654. Schools were selected to provide a suitable spread in terms of the following stratification criteria: geographical distribution; rural or urban; size; male/female/mixed students; management type; school language; DEIS status. Schools were allocated to one of the three subjects, two of the subjects or all three subjects.

The participating schools and the subject combinations they were allocated are listed in Table 2.1 below.

Table 2.1: Participating schools

Dublin				
Blackrock College	Dublin			physics
Scoil Chaitríona, BAC 9	Dublin		chemistry	
Dominican College, D 4	Dublin	biology	chemistry	physics
Fingal Community College	Dublin			physics
St Tiernan's Community School	Dublin	biology		
Malahide Community School	Dublin		chemistry	
Cork & Waterford				
St Francis College, Rochestown	Cork	biology		physics
Coláiste Daibhéid	Cork			physics
Coláiste Pobail Naomh Mhuire	Cork	biology	chemistry	
Kinsale Community School	Cork		chemistry	physics
St Augustine's College	Waterford		chemistry	physics
Rest of Munster				
Pobailscoil Chorca Dhuibhne	Kerry	biology		
Ardscoil Rís	Limerick			physics
Castletroy College	Limerick			physics
Presentation Sec School, Clonmel	Tipperary	biology	chemistry	
St Patrick's Comprehensive	Clare	biology	chemistry	
South Leinster				
Patrician Sec School, Newbridge	Kildare	biology	chemistry	physics
Castlecomer CS	Kilkenny	biology	chemistry	physics
St Peter's College	Wexford			physics
Blessington CC	Wicklow	biology		
Tullamore College	Offaly	biology		physics
Ulster & North Leinster				
Royal & Prior Comp School	Donegal	biology	chemistry	
Athlone Community College	Westmeath		chemistry	physics
St Olivers CC	Louth		chemistry	
St Aidan's Comp School	Cavan	biology		physics
Eureka Sec School	Meath		chemistry	
Connacht				
Holy Rosary College, Mountbellew	Galway	biology		physics
Sacred Heart School, Westport	Mayo		chemistry	
Sligo Grammar School	Sligo	biology		
Roscommon Community College	Roscommon	biology	chemistry	physics

At each stage of the trial, it was impressed on schools and their students that participation in the trial was voluntary for students. Students over the age of eighteen were allowed withdraw of their own choice and students under that age were allowed withdraw with parental consent. By the time the trial took place, the number of students involved had dropped to about 1100.

2.2.4 Preparation at school level

Before the trial, schools were asked to submit assessments (high, moderate, or low) of each student's competence at practical work in the subject(s) that they were taking in the trial, as judged by their teachers. Schools were also asked to submit the percentage score that students achieved in the trial subject(s) in their schools' house examinations in summer 2017. This information was requested so that a measure of each student's practical skills in the subjects and their general ability in the subject would be available to compare with the results produced in the trial as part of the technical assessment of the validity of the practical examination.

Schools were requested to generate a trial ID number for each of their students for use in the trial in accordance with instructions sent by the trial team. The first two digits of a student's trial ID identified the trial school and the final three digits identified the student. The information linking trial ID numbers to individual students was retained at school level and not forwarded to the SEC or provided to examiners.

Before the trial examination, schools received two sample tasks for each subject they were involved with. They also received equipment lists for each subject being trialled in the school. Each equipment list covered all of the requirements for the tasks in the subject. Schools were asked to prepare the laboratories, equipment and consumables for use in the trial.

In the week before the trial, schools were requested to assign students to examination sessions and, following this, they were contacted by the examiner assigned to the school with the date(s) for the trial practical examination(s).

2.2.5 Selection and training of examiners

The trial practical examination had two aspects, the assessment of *practical skills* through direct observation by the examiner, and the assessment of *results and analysis* through the work completed in the task booklets. For examiners, the first of these aspects was an entirely new form of assessment, while the second involved – procedurally at least – a more familiar form. For practical reasons, it was decided that teachers to be selected as examiners should be drawn from the pool of those who already had examining experience in a science subject in the Leaving Certificate or the Junior Certificate examination, or who were teaching in a trial school. Teachers from the schools that had been selected to participate in the trial were invited to apply to examine in the trial, as were the teachers who had tried out the tasks in April 2017. Teachers who were employed as examiners for relevant state examinations in summer 2017 were also invited to apply.

Application forms and outline details of the role of examiner were compiled by the trial team and sent to potential examiners for return by 5 September 2017. Confirmation was required on the application form that the applicant's school was prepared to release them to act as examiners for the trial (with substitution costs recoverable). The selection of examiners for the trial was carried out by the EAMs on Friday 8 September. Examiners were selected as follows: Biology: 16; Chemistry: 9; Physics: 7. Five advising examiners were appointed for Biology and three each for Physics and for Chemistry. This ratio of advising examiners to examiners is higher than is usual for practical examinations. This was due to the particular needs of the project and to provide for a sufficient amount of double-marking

to facilitate a technical analysis of marking reliability. It would not be required in the longer term in the event of a full rollout. In all, 43 examiners were appointed.

Training of the examiners and the advising examiners took place on Friday 6 October and Saturday 7 October. As well as informing the examiners about the trial and the tasks that had been set for the trial, examiners were trained in the direct assessment of students' practical skills as they performed the tasks. This was a new endeavour for examiners and EAMs.

2.2.6 Weather disruption

During the week before the trial, each examiner contacted his/her assigned school(s) and arranged the date(s) for the practical examination. Unfortunately, the planned window for the trial coincided with Hurricane Ophelia. As schools were required to close for the first two days that had been arranged for the practical examination, many of the agreed dates had to be rearranged.

Despite the severe disruption to school life caused by this weather event, all of the trial practical examination sessions took place and the trial was successfully implemented. This success under the circumstances was due to the commitment, dedication and resourcefulness of all involved, including school management (of both trial schools and those releasing examiners), teachers involved in the trial, examiners, students, EAMs and SEC administrative staff.

2.3 Carrying out the trial in schools

2.3.1 Session details

All planned trial sessions were carried out successfully in schools and were completed by Friday 27 October. Table 2.2 below gives the relevant details for each subject.

Table 2.2: Summary of examination sessions in practical assessment trial

	Biology	Chemistry	Physics
Total number of students	523	288	294
Number of Irish-medium students	19	3	7
Number of schools	16	16	17
Number of examination sessions	64	34	38
Range of session size	4 – 12	3 – 12	5 – 12
Average session size	8.2	8.5	8.8

The nature and conduct of the practical examination sessions, including the marking, is described in more detail in the next chapter.

2.3.2 Monitoring and observation

Advising examiners monitored a total of 36 sessions across the three subjects. As well as the supervisory and monitoring work that is usual in such practical examination sessions, the advising examiners also marked the students' practical skills, using the same marking

scheme as the examiner, but each working independently of the other. Because of this, the trial generated two independently generated marks of the same performance in 271 cases.

The EAMs and the members of the implementation group visited a number of sessions to observe the conduct of the trial, as did some members of the steering group. The external project reviewer observed the conduct of sessions in two schools and spoke to some teachers and students involved in the sessions observed.

2.3.3 Filming

The SEC arranged for the video recording of one practical examination session in each of the three subjects. This was to create a visual record of what happened during the trial practical examination sessions, to facilitate informed discussion following the trial. The full-length video recordings and shorter edited versions can be accessed through the links below.

Full-length versions

Leaving Certificate Biology – Practical Assessment Trial
https://vimeo.com/242052353/f2d715cbc0

Leaving Certificate Chemistry – Practical Assessment Trial
https://vimeo.com/242049015/225bc3a00a

Leaving Certificate Physics – Practical Assessment Trial
https://vimeo.com/242216667/969d25f78e

Edited versions

Practical Assessment Trial – BIOLOGY – Short Video 10min 48 sec
https://vimeo.com/247182259/17a410901b

Practical Assessment Trial – CHEMISTRY – short video 12 mins
https://vimeo.com/247181195/e731845888

Practical Assessment Trial – PHYSICS – short video 9 mins
https://vimeo.com/247183151/1be249d39e

2.4 After the trial – feedback, marking, and data-capture

2.4.1 Marking of task booklets

As previously noted, assessment took two forms – the examiner awarded half of the marks for the *practical skills* observed during the session. The remaining half of the marks were to be awarded later for *results and analysis* (the work done in the booklets).

Marking of such booklets is a form of assessment with which the SEC already has considerable experience. There are well-established procedures for handling both the logistics and the quality assurance of this operation.

For the purposes of the trialling, the examining team marking the booklets consisted of the same team of examiners who carried out the sessions in the schools. However, these were two distinct operations – examiners were not assigned the booklets for the same schools

where they had examined, and accordingly were not attempting to link what was written in a given booklet with what that candidate had been observed doing on the day. In the case of a full rollout, it is likely that the booklets would be marked by a different team, such as the team marking the written papers.

On 2 and 3 November, marking conferences were held at which the marking schemes for the task booklets were agreed, applied to samples of work, discussed and finalised. Examiners then took home their allocation of booklets and marked them over the following weeks. Marking was carried out using the same procedures as for the SEC written examinations. Each examiner applied the marking scheme agreed at the conference and the work of each examiner was monitored by an advising examiner during the marking. The marking continued until about 22 November, at which point examiners returned their mark sheets and the results were recorded in the science trial databases.

2.4.2 Mark databases

A database for analysis of marks and related data was prepared for each of the three subjects. In the case of each student, the record included, where available, the trial ID number, the session number, the task number, the mark awarded by the examiner under each subheading of the marking scheme for the practical skills element, the marks awarded by the advising examiner (if the session was monitored), the moderated mark following application of monitoring and moderation procedures, the total mark awarded for the assessment of results and analysis (the booklet), and the data supplied in advance by the school – percentage mark in preceding summer test in the subject, and teacher assessment of practical skill level (high, moderate, low).

These databases were used for the statistical analyses that are reported on in chapters 5 and 6 .

2.4.3 Capture and analysis of feedback (questionnaires)

Following the trial, questionnaires were issued to students, examiners, teachers in the trial schools, and principals. Of the 30 schools in the main trial, 27 returned principal questionnaires, 29 returned student questionnaires, and 29 returned teacher questionnaires. Of the 43 examiners and advising examiners involved in the main trial, 40 returned examiner questionnaires.

Analysis of the questionnaires was contracted out to an agent with specialised expertise in this area. The outcomes of the analysis form the basis for Chapter 4 of this report.

2.4.4 Other information collected

Other data was collected during the trial from the examiners as follows: mark sheets for assessment of practical skills and for assessment of results and analysis; examiner reports on both of these elements; and reports on the practical examination centres. During the marking conferences on 2 and 3 November, examiners' observations and opinions on their experience of the direct assessment of practical skills were sought in a discussion forum and were recorded.

Examination centres were visited during the trial by members of the Implementation Group and any reports they made have been taken account of in the preparation of this report.

Chapter 3 The practical assessment in detail

This chapter describes how the practical examination was carried out, with particular reference to the element involving the assessment of practical skills by direct observation. The procedures for examining students' written work in the task booklets and recording all of students' marks are also outlined.

3.1 The practical examination sessions

As the proposed practical examination is of 90 minutes' duration, the trial was organised into 90-minute 'sessions'. Within each subject, up to three sessions per day were arranged in a laboratory. Each session involved up to twelve students, working individually and (in general) on different tasks, observed by one examiner. Thus, each laboratory could conceivably have accommodated up to 36 students in a day. The examiner awarded marks (out of 60) for the practical performance. The work completed by students in the task booklets during the practical was collected and awarded marks (out of 60) later.

The tasks for each of the subjects were based on the mandatory experiments in the current syllabus. Fourteen task booklets were set for each of the three subjects. Two of these were issued to trial schools as sample tasks to help the students prepare for the trial, and the remaining twelve were used in the trial. A separate sheet, common to all tasks and subjects, entitled *Instructions and Information for the Practical Examination*, was issued to each student in the trial, and a common format and layout was used for the booklets in each subject. Further information on the content of the tasks is contained in Chapter 6.

There were slight differences across the three subjects in the presentation of tasks and booklets. In Biology, nine of the twelve trial tasks consisted of a single activity, while three were divided into two sub-tasks each. Each task and sub-task had a title stating the nature of the investigation involved, a short introductory paragraph, followed by a list of instructions, followed again by a set of questions with spaces for answers. In Chemistry, each task consisted of two sub-tasks. Each sub-task had a title that stated what was to be done in that sub-task followed by a list of instructions with questions and spaces for answers interspersed with the instructions. In Physics, each task had a title that stated what was to be done in that task followed by a list of instructions with questions and spaces for answers interspersed with the instructions.

Students could ask for help if they needed it, but the examiners were instructed not to question the students directly. 'Pseudodata' was available for students who were not able to generate their own data – for example as a result of a spillage, a chemical that was improperly prepared, or a procedural error by the student. In certain cases, obtaining help or seeking to use the available pseudodata was penalised, depending on the circumstances. The task booklets were used by the students to record observations, data, and data analyses, and to answer questions during and/or after their experimental work. Graphs were drawn on standard SEC graph paper. Tidying up the workstations at the end of the practical examination session was part of each task.

3.2 Organisational procedure

Before the first session in each school, the examiner held a briefing meeting with students during which the examination procedures and rules were explained. Before each individual session, students were invited to enter the examination centre about ten minutes before the session start time. The tasks had been randomly assigned by the examiner to the students in advance, and these task allocations were listed on the *Direct Assessment Mark Sheet*. As far as possible, each student was allocated a different task[2]. Students were directed to the numbered workstations corresponding to their tasks and given pre-printed labels with their Trial ID number to wear during the examination session.

Students were reminded to put on appropriate safety clothing and that mobile phones were not permitted in the examination centre. They were asked to read the *Instructions and Information* sheet on their workstations. Students were reminded that they were allowed to ask for help, by putting up their hand, and that such help might involve losing some marks, depending on the circumstances and the kind of help sought.

The examiner distributed the task booklets. At the appropriate starting time, students were given permission to start and reminded that the first ten minutes of the examination session was to be used only for reading the task booklet, preparation and planning. They were subsequently informed when this ten-minute period was up so that they could begin assembling apparatus. They were told that they would be alerted again ten minutes from the end of the session, and that the last five minutes would be for clearing up.

During the first ten minutes, students were permitted to gather apparatus, but they were not permitted to assemble it or start working on the task or completing the booklet. During the next 75 minutes, students assembled the appropriate equipment, manipulated resources, carried out experimental activities, and recorded observations and measurements.

The examiner used a clipboard and the *Direct Assessment Mark Sheet* to assign marks or make notes discreetly as he/she moved around the laboratory, observing the students carrying out their tasks. The examiner assessed each student's practical skills under the five assessment categories given in Section 3.3 below.

The only types of communication allowed between an examiner and a student were:

- exchanges relating to a request for help by a student
- a student calling the examiner to observe a particular activity, as required in some tasks or as requested by the examiner
- communication necessitated by any situation that threatened the safety of the student or of other students or that compromised the efficient running of the examination centre.

[2] As the trial was being led comparatively early in the school year, there were some tasks that could not be given to students in a particular school, as they had not yet covered the relevant sections of the course. The necessary adjustments were managed locally between the examiner and the class teacher.

Where a student needed help that involved a penalty, the help required was given to enable the student to progress. The penalty was recorded in the appropriate place on the mark sheet and a corresponding deduction of marks made.

As the session progressed, the examiner had opportunities to observe the students' selection of apparatus, chemicals and other materials, assembly of apparatus, use of apparatus, and safe and efficient way of working in the laboratory, as described in mark categories 1, 2, 3, and 5 in Section 3.3 below. Towards the end of the session, the examiner awarded marks for each of these four categories to each student, based on the student's overall 90-minute performance. For some specific tasks, there were relatively few opportunities for the examiner to observe a student making a measurement or recording an observation (mark category 4). The examiner may have asked a student doing such a task to call him or her to observe key moments, or the examiner may have asked students to repeat a key step involving measurement or observation. Marks for this category were nonetheless awarded to each student based on the student's overall 90-minute performance, and penalties were deducted where help was given or data provided. The lowest possible mark in this category, even when penalties applied, was zero. That is, a deduction for receiving help or receiving pseudodata could not result in a negative mark being awarded.

Students were alerted when there were ten minutes remaining in the session. Before the last five minutes of the session, they were alerted again and reminded that the last five minutes were designated for tidying up. At the end of the examination session, the examiner instructed students to leave their task booklets and any graph paper used at their workstations. As mark category 5 included the marks for tidying up, the mark for this category was not finalised for any student until that student was finished tidying up. After all students had left, the examiner completed his or her session mark-sheet and collected the session task booklets, which were later sent to the SEC for marking.

3.3 Assessment objectives for assessment of practical skills

The assessment objectives for the directly observed assessment of practical skills were organised under five headings and the criteria for awarding marks were as follows.

1. Selection of apparatus, chemicals & other materials
- apparatus suitable for task
- sufficient apparatus appropriate to task
- chemicals/other materials needed for task

2. Assembly of apparatus
- correct assembly
- manipulative skills in assembly

3. Use of apparatus
- candidate carries out task as directed
- manipulation of apparatus during conduct of task
- co-ordination and dexterity in the use of equipment
- apparatus used appropriately

4. Observations/measurements
- correct observation/measurement technique
- accurate observations/measurements
- sufficient repetition where appropriate

5. Working safely & efficiently & cleaning up

- personal safety and safety of others
- economic and safe use of resources
- tidy work practices
- task completed within the given time
- adherence to safe work practices
- cleaning of work area

Each heading carried up to 12 marks and the marks that could be assigned by the examiner for each were 12, 8, 4 or 0, according to the marking key given in Figure 3.1 below.

Figure 3.1 – *practical skills* marking key

MARKING KEY FOR DIRECTLY OBSERVED PRACTICAL SKILLS			
High level of achievement	**12**	Low level of achievement	**4**
Moderate level of achievement	**8**	Not achieved	**0**

Chapter 4 Feedback from participants

This chapter examines the trial from the perspective of its participants – those in the schools and the examiners. While the students in the thirty schools and the examiners were the principal participants in the trial of practical assessment, they were supported by the teachers and management of the schools. All of the trial participants – the management, students and teachers in each of the schools, and the examiners, were facilitated in giving their views of their experience of the trial and, arising from that experience, their opinions on such important issues as the impacts of practical assessment in schools and how practical assessment can be organised. The material in this chapter is based on the responses to questionnaires that were sent to schools and examiners following the trial. The openness of the respondents and their attention to detail in their responses are the foundations of this chapter. Where more than one view on an issue was expressed by the respondents, the chapter gives the more common view first and/or tries to quantify whether a response was given by most, several or a few respondents. In order to give a good understanding of the detail of the practical examination, several quotes are included from responses in the questionnaires.

This chapter opens with the analysis of survey responses from teachers who were involved in the trial and examines the impact of the trial on their schools. The teachers' opinions on potential impact of such a practical assessment if it were to be implemented in the Leaving Certificate examination are reported. Their opinions regarding the organisation of such a practical assessment conclude this section.

In the subsequent sections, the trial is looked at from the point of view of the principals of the schools that took part in the trial and of the students who were participants in the trial practical examination. In the case of the principals, the impact of the trial on the schools, and the future potential impact of practical assessment are examined. The section dealing with the perspective of the students looks at their opinions on the fairness and appropriateness of the practical examination and on how difficult or otherwise they found the tasks. It also examined how they experienced the trial.

The remainder of the chapter reviews the experience of the examiners in the trial. As three trial examinations, in Biology, in Chemistry and in Physics, were being run simultaneously, some parts of the examiners' responses are examined separately for each subject. This part of the chapter provides considerable detail on what happened during the direct assessment part of the trial practical examination. It also includes examiners' assessments of the trial examination and of how difficulties might be addressed.

4.1 Schools

Principals and teachers in the trial schools were asked to give information on how their schools prepared for the trial examination and the impact that the trial had on their schools. They were both asked about potential future examiners from their schools and about the implementation of external practical assessment in the Leaving Certificate examination.

4.1.1 Teachers

Three categories of teacher were involved in the trial and teachers in each category responded to specific parts of the questionnaire. The role of *subject teachers* whose students participated in the trial was to familiarise the student with what would happen in the trial and to prepare them for the experience. To assist them in this role, SEC sent sample examination material to each school about six weeks before the trial commenced. Many subject teachers helped in preparing laboratories for the trial who were not designated support teachers (DSTs).

Each school had a *school contact teacher* who carried out the role of contact between SEC and the school. This involved all communication from the SEC trial team to the school being copied to them by email for circulation within the school. They ensured that the teachers who were involved in the trial received all relevant communication.

The category of teacher with the greatest responsibility in relation to running the trial was the *designated support teacher* (DST). These teachers supported the preparation of the trial in the school and the running of the trial. The designated support teacher for each subject in a trial school was responsible for ensuring that the examination centre for the trial examination was properly prepared, including preparation of consumables such as chemical solutions, and was on-call during the examination to assist in the event of equipment failures, spillages, breakages, shortages or emergencies. They also briefed the examiners about the school and worked with the examiners when they were in the school.

The extent of the involvement and commitment of teachers in the trial is evident in their responses. Also evident is the level of co-operation that existed among the teachers and between the teachers and the examiners. Without the commitment and effort of the teachers, the trial would not have succeeded in its aims.

There was a good response from schools with regard to returning the teachers' questionnaires. In all, 73 completed teacher questionnaires were returned from 29 schools. The subject teacher section of the questionnaire was completed by 63 subject teachers, the school contact teacher section by 14 school contact teachers, and the designated support teacher section by 41 designated support teachers.

4.1.1.1 Communication between SEC and schools

School contact teachers considered that email followed up with written letters and phone calls to have been a most effective means of communication with schools from SEC. They considered that communication with the school could have been more effective if more details about how the trial would transact in their school had been issued. In particular, contact teachers considered that information about the specific list of experiments to be used in the trial could have been given to schools.

According to the contact teachers, email and face-to-face conversations or meetings were the principal ways in which communication about the trial took place.

Advice from the school contact teachers to the SEC on improving communications with schools included the use of follow-up emails and making direct contact with the teachers whose students were involved in the practical examination. The teachers were generally satisfied with the level of communication during the trial, but they stated that if there were

to be a national roll out, then SEC should maintain a similar level of communication with schools as during the trial. A separate contact teacher for each subject was suggested instead of a single contact teacher.

4.1.1.2 Impact of the trial on the school

For many teachers, the preparatory work for the trial examination meant that they missed some of their timetabled classes. In some cases, some or all of the preparation work was undertaken in the teachers' own time. As well as many subject teachers missing some classes while the trial was in progress, laboratories were unavailable for a period because of the trial. For many subject teachers, the preparation time was considerable, and in some cases well in excess of what the teachers had expected. Teachers spent time

- preparing the students in terms of their learning and skills development
- preparing the laboratory and preparing the equipment
- preparing and labelling the chemicals/materials.

> I lost teaching time with my sixth year students to brief them on all aspects of the trial and I also lost teaching time with my fifth year & third year students.

> I need 3 hours to prepare some of the solutions etc required. I had to leave my classes work to do during that time.

> All of my other free class periods and after school were taken up with organising the practical a week before. Also had to be in at 8 in the morning of the practicals as the practical took place at 9.

> The students were very worried about it and wanted extra classes to do revision this took from the already huge course that we have to cover.

> Had to leave work for classes that were supervised.

> Prep for trial, particularly chemistry, was very onerous and used up a lot of free time...as a result, planning and prep for other lessons suffered. It the trial format was adopted as the official exam I would give serious consideration to not teaching chemistry at LC level any more.

> As with all exams, rote learning will be necessary for the students and rote practice doing the practicals was required. This allowed less time in class for discussing the amazing aspects of Physics not on the course.

> My lab was used for the duration of the trial, this required my classes to be conducted in general classes. As I was also the DST it meant that I was not available to teach these classes, SO class time lost.

> The trial forced us to do an overdue stocktake

> Far too much time was lost in the weeks building up to the practical due to the VAST amount of chemicals and solutions DEMANDED.

Suggestions made by about one quarter of the respondent subject teachers to reduce the disruption to teaching included the provision of laboratory technicians to help in preparing the laboratory and the chemicals and materials needed for the trials. Approximately 10% of the teachers responding also suggested that they be allowed time to prepare for the practical examination, perhaps as free class periods. Other suggestions included the provision of an option for schools to hold the practical examinations outside of class time and to remunerate the parties involved. Some teachers felt that sufficient advance notice of the topics or a limit on the number of experiments to be examined would give them enough time to prepare students. Other suggestions, each made by about 5% of the respondents, were to update laboratory equipment, to provide more laboratory space, and the examiner to bring the required solutions with them to the examination.

Most (about 57%) subject teachers agreed that the equipment available in their schools was sufficient for the trial. Areas where there were shortages were lack of laboratory coats, glassware equipment such as stopwatches, chemicals and biological materials (in particular *Elodea*). More than 80% also considered their laboratories suitable for the conduct of the practical trial. However, the remaining teachers said the practical examination put some pressure on their laboratories due to their size and/or the demands of access to them.

4.1.1.3 Impact of practical assessment on teaching and learning

For most subject teachers the principal impact they would expect of having practical assessment in the Leaving Certificate examination would be to have more student practical work in classes and to facilitate more individual student practical work. Most would include practical assessment as part of their house exams. There was a consensus among the teachers that with practical assessment they would place a greater emphasis on teaching their students the techniques of independent hands-on work, including setting up and conducting experiments, and clearing up equipment and materials.

However, one quarter of the respondents felt that such a practical approach, including an emphasis on individual practical work, would not be possible in the class time available. They would not have sufficient time to prepare for a practical examination and at the same time cover the practical and theory work on the syllabus. As a result, they feared that students would suffer through not being given sufficient attention. Many teachers felt that while pupils already engage in practical work in the science subjects, facilitating individual practical work would be difficult.

> I already carry out practicals with students but would have to try to get them to work individually which would be very difficult due to lab access and availability of resources

> I would let them be more independent when doing practical work.

> At the moment students carry out the practical work as a preparation for the written examination. I set a time limit on the work and occasionally I may supply data where an experiment takes too long. Also having taught say how to calculate acceleration using a traditional methods I would use a quicker method (data logging) to find the acceleration for Newton's second law. For the practical assessment, this would not have attracted full marks.

> We would use worksheets like the tasks booklets when completing practical work.

> Students are used to doing group work only for practicals.

> Not always possible for students to work individually during practical sessions because of space and equipment shortages. It is mainly carried out in groups.

> I will be using practical work with my current 5th years in their summer exam.

> Due to the fact we have 3 labs in school and 7/8 teachers teaching...there is just no way that we can close down the labs for days to do practical exams.

> Practicals are all mandatory as it is, so no change in terms of teaching.

4.1.1.4 Impact of practical assessment on students

Subject teachers' responses on the impact that having practical assessment would have on their students were overwhelmingly positive. Beneficial effects of practical assessment in Biology, Chemistry and Physics on students' attitudes to and engagement would include

- encouraging students to take responsibility for their learning (>50%)
- focusing students' attention and sense of enjoyment through practical work
- developing students' practical skills and capacity for inquiry
- helping them at third level.

> Attitude & engagement would improve, the practical assessment is very accessible to all abilities, would be a positive addition.

> Students would focus more on completing work/practicals with precision and accuracy if marks were allocated for LC exam.

> More focus on practical work from the beginning of 5th year. This would prepare students better for 3rd level science.

> They would be more focused on the whys in the experiment. Wanting to get a deeper knowledge.

> From what I saw it really increased their enjoyment, they really enjoyed doing the practical, it provoked more questioning than a normal experiment class.

A very few teachers (6%), however, thought that practical assessment could potentially increase the stress and pressure experienced by students and could discourage students from studying the subject(s) because of frustration or irritation with practical work.

> Some students get irritated with practical work as it is. Having a practical exam to do will certainly distance them further from the subject. Over the 30 years I have been teaching Physics, I have seen an increasing awareness by students, of the trivial nature of performing "regulation" practicals. Adding an exam to this is not going to help.

Following the trial, while many (48%) subject teachers did not notice a change in students' attitude, others (33%) did notice some positive changes including

- increased student confidence

- improved recall of experiments and subject content
- increased student questioning and discussion
- an interest and enjoyment in the subject.

4.1.1.5 The implementation of practical assessment

Teachers were questioned about how practical assessment would be implemented. A key consideration, in the context of the external model that is being considered for practical assessment, is the availability of examiners. Other aspects are the proportion of the examination marks to be assigned to practical assessment and when should it take place each year.

Examiners for Practical Assessment

Most (68%) subject teachers said that they would apply to be an examiner for a Leaving Certificate practical examination. Among the reasons they gave were obtaining an insight into the practical examination process regarding organisation, preparation, and standards expected. Improving their teaching was another reason given. Where teachers did not wish to become examiners they said that they would feel uncomfortable and apprehensive about missing time with their classes. The current reported lack of substitute cover in their subject areas was another reason given.

Proportion of marks for practical assessment

While most (57%) teachers considered the 30% weighting for the practical assessment to be appropriate, some others suggested alternatives ranging between 10% and 25%. Most (55%) teachers considered the equal division of marks between direct assessment and indirect assessment in the practical assessment to be appropriate. Most of the minority of teachers who considered the division inappropriate, (25% of the respondents,) suggested that the practical component should be awarded a greater proportion of the marks.

Time in the school year for the practical examination

Teachers gave a lot of feedback about the difficulty of finding a time in the school calendar for the practical assessment that did not clash with other school activities ('mock' examinations, language orals, LC practical examinations in other subjects, new Junior Cycle CBAs, etc) and which was also sufficiently late in the two-year senior-cycle programme to have good course completion. A number of the teachers' suggestions fell between the February midterm break and after the Easter holidays with some also favouring holding the practical examination before the Christmas break.

4.1.1.6 Supporting practical assessment in the school – the designated support teacher (DST)

The designated support teachers were responsible for preparing the school for the trial, liaising with the examiner(s) before and during the trial, and coordinating the restoration of the school laboratories after the trial. They were of central importance to preparing for and running the trial in a school.

Preparing for the trial

All of the DSTs were teachers of the subjects for which they acted as DST and almost all agreed that this is necessary. They said that the DST for a subject must have the subject

knowledge, practical skills and understanding of what is entailed in the practical assessment.

While many (71%) DSTs stated that the chemicals and other materials needed for the trial were no different from those required for the teaching of the subject, some (29%) DSTs disagreed. Most comments from DSTs stated that they had to order additional or replacement stock (equipment, glassware, apparatus, fresh ingredients such as *Elodea*, and chemicals) in advance. In one school, the DST said that equipment and materials were borrowed from neighbouring schools. The Boyle's Law experiment was 'removed' from the list of tasks by another DST because of equipment breakage or equipment shortage.

Most (78%) DSTs said that their laboratory could accommodate up to 12 workstations. Where this was not the case, it was due to inadequate space and too little equipment. DSTs said that getting ready for the practical examination was time consuming, as they had to identify and gather the necessary equipment and preparing the chemicals was not a straightforward task. While DSTs gave preparation times for the trial ranging from one or two hours to 'every minute of spare time for 3 months', typical responses were between 2 hours and 4/5 hours (12 DSTs), between 10 hours and 20+ hours (13 DSTs), and between 1 day to 3 days (8 DSTs). It is clear from the responses of the DSTs that preparation for the Chemistry trial was far more onerous than for the other two subjects.

The DSTs were asked whether there was any disruption of their timetabled classes because they were occupied with preparing for the trial. While most (76% or 31) of the DSTs who responded said that their classes remained uninterrupted because they prepared for the trial in their own time, some DSTs (29% or 12) stated that their classes were disrupted because they were occupied with preparing equipment, chemicals, and materials for the trial.

When asked whether there was disruption of timetabled classes in their subject because the laboratory was being prepared 13 DSTs said that classes were disrupted because the laboratory was unavailable while it was being prepared for the trial. Twenty-five DSTs said that there was no class disruption due to laboratory unavailability before the trial.

During the Trial

Most DSTs missed teaching time while the trial was in progress with the amount of time missed mostly ranging between 1 and 2 days (11 DSTs), between 4 and 6 hours (9 DSTs), and approximately 2 hours (8 DSTs). Six DSTs said that they missed no teaching time or very little time. Most DSTs felt that it was necessary for them to be available when the practical sessions were in progress to help with equipment and to ensure the efficient running of the examination. Supervision and substitution were the principal ways through which teachers' absences from their classes were covered although in some cases teachers continued taking their classes during the trial.

> Not covered, I stepped out of a class to assist in the practical set up. Yes, it was important that I was available.

> Was teaching in room next door. Placed a different TY students each class in the lab…every 40 minutes this student changed but if anything was needed they came to get me.

> Covered by sub teachers. Important to have time to organise the students for next session, be available in case equipment was needed, also to be free to fix room between sessions.

The length of time that was available between practical examination sessions, in schools where multiple sessions were involved, mostly ranged from 40 minutes to 1.5 hours (in 18 schools) and most DSTs considered this to have been sufficient.

After the Trial

Fifteen of the DSTs said that it took between two and three hours to restore their laboratories following the trial and much of this work took place outside school or timetabled hours.

Reflections

The DSTs were asked to reflect on the number of practical session held each day, the fairness of the examination, and their role.

Most (73%) DSTs felt that two or three practical examination sessions could be held each day with between one and two hours required between sessions. While many DSTs considered the practical examination to have been fair to students, others felt that the variations in tasks in terms of level of difficulty or time required and/or the lack of choice in tasks to have been somewhat unfair.

DSTs considered the following to have been the most satisfying part of their role

- working with the students and facilitating them to develop and showcase their skills
- hearing feedback, positive and constructive, from the students
- learning what the students did not know
- observing the practical examination process in action.

> Seeing the students' confidence after carrying it out, they felt very important and respected! It was great to get an insight into how it may develop in the future

> Seeing increased importance placed on my subject. Getting to meet other teacher(s) who acted as examiner(s) and learning from them

The most difficult part of the DST role was considered by many to be preparing for the trial and setting up in terms of sourcing materials, chemicals, and equipment and preparing for the implementation of the trial.

> To try to encourage students to participate in the trial; the responsibility that comes with the role, time consuming; losing valuable teaching time with my students

> Making sure I left all equipment and solutions clearly laid out for students to find

Some DSTs considered that the preparation work was the most time-consuming aspect of their role as DST.

> Checking all the equipment in August then gathering them. NB: Our school is an Irish-speaking school and the students were only given the Irish version. In every state exam they are obliged to have both Irish and English. Some students found the Irish hard to follow and therefore found it hard to answer the questions.

> Making up solutions, chemicals ready. We have two labs for six teachers so it is very difficult to ensure that equipment is in order and that you won't run out of chemicals. I feel if this is to go ahead that the experiment list should be shortened for the exam--maybe 8 different experiments listed and maybe, the students could have a 10 minute oral to discuss what they have written up in their practical copy (as I did in college).

It should be noted that while teacher substitution was allowed and was paid in respect of the trial, much of the work carried out by teachers was in their own time. While this is an indicator of teachers' commitment to the trial, if practical assessment were to be included in the Leaving Certificate examination, this issue needs to be considered further.

4.1.2 Principals

The principals of 26 of the 30 participating schools returned questionnaires. The principals' questionnaire sought information on the impact of the trial on their school and their opinions on implementing practical assessment in the Leaving Certificate examination. The material in the sections below is based on their responses.

Impact of the trial on the school

The principals of the participating schools were asked about how their school had been affected by the trial. They were positive about the possibility of practical assessment as part of the Leaving Certificate examination and most felt that it was a positive and necessary change to the assessment of second-level students.

According to the principals, there were three main sources of disruption in preparing for and running the trial. These were

- restrictions of access for students and teachers to laboratories
- release of teachers from their usual teaching duties and the resulting loss of teaching time for subjects other than the trial subjects
- during the trial practical examination students involved in the trial were absent from classes in other subjects.

In their responses, principals stated that the time required by their schools to prepare for the examination ranged from 4 hours to 30 hours. Some principals referred to teachers working outside school hours to prepare for the trial. All principals referred to the difficulty in getting suitably qualified substitute teachers to cover classes for teachers involved in preparing for and running the science trial.

> Suitably qualified subs hard to get -depends on timing of the trial (time of year) / 'it is virtually impossible to get substitute teachers.

Some principals suggested that the work of preparing for practical assessment should be shared among all science teachers including teachers whose classes would not be undergoing practical examination. The example was given of all the science teachers being involved in laboratory maintenance, including stocktaking. Another principal referred to the desirability of having more than one support teacher for Biology, because of the large number of students involved.

Resources for practical assessment

Principals were asked whether their current resources would be adequate if practical assessment were to be included in the Leaving Certificate examination. Almost all principals noted that funding would be required for materials, chemicals, solutions, equipment, and laboratories, including for increased quantities of breakables and renewables because of increasing individual student practical work. Laboratory storage facilities in many cases were said to be inadequate and not fit for purpose.

The need for school laboratory technicians to oversee the running of the laboratory and contribute to the preparation and facilitation of the practical examinations was referred to by several principals. In one school where there was a laboratory technician, the principal stated that without a laboratory technician 'we would have found it very difficult to prepare the labs and materials needed for the practical exam'.

> Inadequate resources: no. of rooms. Increase in individual student work will increase demand for breakables and consumables.

Examiner supply

Principals, while favourably disposed to releasing teachers to act as examiners in a practical examination, referred to the lack of suitably qualified substitute teachers as a factor that would prevent this. Six out of the nine principals who responded said that they would release teachers. One of them stated that this would be contingent on the availability of substitution.

> Yes most definitely, but only if qualified science teachers are available to continue to teach the classes while the science teacher is away from the school examining.

Principals saw the benefits to teaching, learning and assessment of their teachers becoming examiners.

> Teachers would gain valuable experience and knowledge regarding the practical exam which could be disseminated to other teachers, and more importantly used to plan classroom lessons.

> Students would benefit from the experience and knowledge the teachers would gain from being examiners.

Principals who would not release teachers referred to the current difficulties with regard to the supply of substitute teachers and the loss of teaching time to their classes.

> reduced contact time, lack of continuity, pressure on teaching with orals and practicals.

Timing of a practical examination

There was no consensus regarding when a Leaving Certificate practical examination should take place. Suggested time for the practical examination ranged from the summer of 5th year, through February of Leaving Certificate year to after the Easter break, or as part of the main examinations. Four principals suggested that the practical assessment should take place outside of teaching time, i.e. on Saturdays or during holidays.

Principals made other suggestions regarding the practical assessment. These included having separate weeks for each subject, the examiner to bring the required materials on

the day, and having an assistant/technician present during the practical examination. One principal suggested that the practical examination could be video recorded and submitted for assessment.

Impact of the practical examination on teachers

Principals were asked what impact they thought practical assessment would have on how their teachers taught. They commented positively on the impact of the trial on their teachers. One principal referred to collaboration and discussions among the teachers and the excitement among the teachers at the prospect of 'being at the forefront of curriculum change'. Another principal said that it 'enhanced cooperation and dialogue on teaching and learning within the department'.

> Brings the long needed practical element to Science which brings the subject to the real world of Science as a practical, experimental endeavour

Principals referred also in their responses to the need for their teachers to receive regular continuous professional development with regard to teaching and assessing students' practical skills in the subjects. Some expressed concern about the extra pressure, additional work, and added stress of implementing practical assessment. They feared the effect on teachers if they had to undertake more practical work with their students, as the school schedule was already full.

Impact of the practical examination on students

Principals in their responses were enthusiastic on the effect that they felt practical assessment would have on their students. They saw it as bringing bookwork to life and allowing students to experience real laboratory work. Most principals stated that practical subjects should have a practical element. Other benefits seen for having practical assessment were embedding in students a better understanding of the subjects. Principals considered that students in the trial were highly motivated, engaged, enthused, more confident and more independent in their learning.

> More practical work means a more student centred approach where students are actively involved and engaged in the subject. This must be a good thing.

> Allocation of valuable exam marks for the skills and practice of doing experiments is critical and long overdue in what are essentially subjects in which experimentation is a key aspect.

> The students deserve credit for these skills. It also prepares them better for the real world of Science after second level.

Along with the overall favourable reaction to the trial practical examination, there were some drawbacks seen. One of these was the common level nature of the practical examination. Principals were concerned about reports from student that the tasks were different in terms of the time required to do them and their inherent levels of demand. One principal expressed a concern about the safety implications of the practical examinations with up to twelve tasks being conducted at the same time, in particular with regard to Chemistry and the 'electricity element' of Physics. Three principals expressed regret that no feedback on their performances was provided to the student participants.

Unfair - common level of provision placed unfair demands on pupils.

Students complained about the unfairness of the different practicals, different lengths/difficulty level.

Imbalance between the different tasks (some being more straightforward than others)

4.1.3 Students

The student questionnaire, which was formatted as in the summary analysis in Appendix D, sought information from students on their reaction to undergoing the trial practical examination in each of the subjects as well as their opinions on the practical examination as a test of their laboratory skills. All schools except one returned the students' questionnaires; 879 in all were returned[3].

Because of the amount of time it would take to analyse all of the student trial data, it was decided to sample the data. The purpose of the questionnaire was to establish the range of issues that students had about the trial practical examination. It was expected that it would also establish their general opinion on the examination. It was considered that a sample of 151 student questionnaires, selected as described in the following paragraph, would be sufficiently representative of the group of students who participated in the trial. It is the intention to carry out a similar analysis on the remaining 728 questionnaires should it be considered necessary.

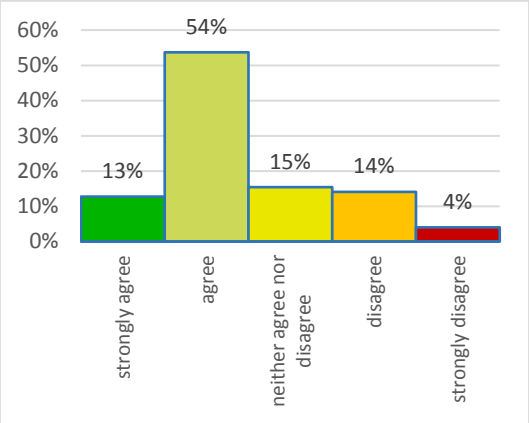
The trial practical examination in Biology took place in 16 schools, Chemistry examinations in 16 schools, and Physics examinations in 17 schools. Where the numbers allowed, three questionnaires were selected at random in respect of the students in each of these subject examinations. In some cases where a large number of students were in a school-subject group, up to five questionnaires were selected at random. The number of questionnaires sampled was 151, 50 each for Biology and for Chemistry (9.6% of the Biology participants and 17.3% of the Chemistry participants) and 51 for Physics (17.3% of the participants).

The first three tables in Appendix D summarise the outcome of the analysis of the student questionnaires for each of the three subjects, and the fourth table is a composite across the three subjects of the responses to the first five questions. These five questions are relatively generic questions about this form assessment, and the responses are illustrated in the Figures 4.1 below. Most of the students (66%) in the sample agreed or strongly agreed that they had enjoyed participating in the trial. 83% of the students considered the trial practical examination to be a good way to test their practical skills, although, interestingly, a much lower percentage (43%) considered it to have been fair. 57% of the students said that practical science skills should be tested as part of the Leaving Certificate, and 79% said that they would pay more attention to practical work in class if there were a practical examination as part of the Leaving Certificate.

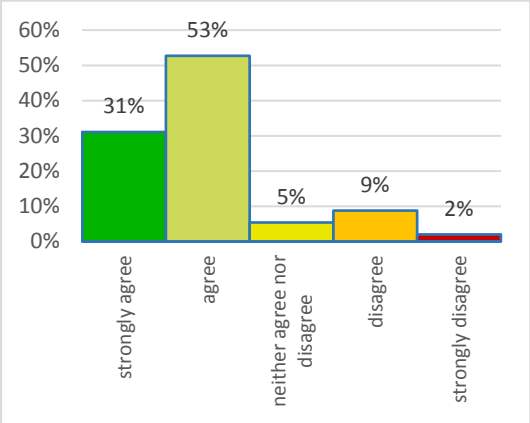
[3] This excludes responses from the three schools in which the digital trial took place.

Figure 4.1 Student responses to questions 1 to 5 on the questionnaire

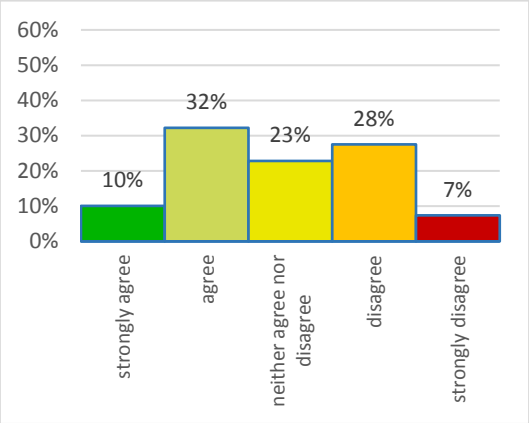
I enjoyed participating in the trial



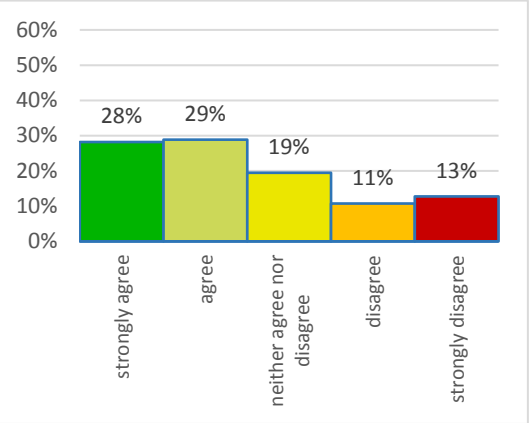
This was a good way to test my practical science skills



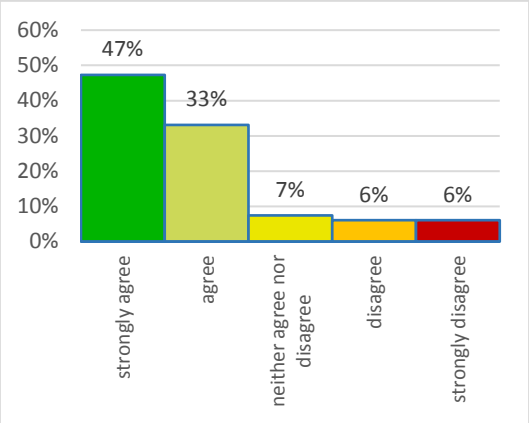
The test was a fair way of assessing different student science skills



Practical science skills should be tested as part of the Leaving Certificate



I would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate.



Further questions related more to the details of implementation of the trial. Most students found the examiner's briefing before the examination to be helpful. It is interesting to see the student view on how busy their examiners were. In the case of Biology 11 out of 49 students agreed or strongly agreed that their examiner was too busy to observe them properly. The figure for Chemistry was 25 out of 50 and for Physics 16 out of 50. This issue may have influenced the students' views of the fairness of the assessment, as might any perception that the tasks were not all of a similar level of difficulty.

The responses to the first five questions

Students' responses to questions 17 to 20 of the questionnaires were broad ranging and in order to best contextualise them they are summarised here by subject.

Biology

Students found setting up the task and gathering the equipment to be the easiest part of their task along with concluding the task and cleaning up.

> Clean up. Plenty of time given and it is not a difficult thing to do.

> Cleaning up afterwards. I have much practice in this area.

> Finding things I needed. It was easy as I knew where everything was.

The hardest parts of their task for many students were finding the equipment and doing the experiment.

> Finding stuff. Couldn't find anything.

> Figuring out how to carry out the experiment. I didn't feel it was clear from the method what I had to do.

> Sharing the equipment with other students.

Other comments students made regarding the Biology trial were that it was well-organised and that there was a perception of inequality due to the differing lengths and difficulty of the tasks.

> I highly recommend it.

> I think that some people had far more difficult tasks to carry out and set up making it unfair in my opinion.

> Would not recommend to roll this trial out as currently is

General comments made by students and not already referred to were in relation to the trial being a good opportunity for practical learners and that it was good to have an external examiner.

> I thought it was fair to be assessed by an external examiner. I think it is a good opportunity for practical learners. I would like if this was part of my leaving cert exam.

> I was happy that the examiner wasn't a teacher in the school. He was someone who didn't know us or our grades and was fair grader. I wouldn't mind having this in the leaving cert and the exam questions I feel it would help.

Chemistry

The easiest parts of the Chemistry practical examination were considered by students to be the setting up of the task, carrying out the task and clearing up.

> Setting up apparatus. I knew where everything was.

The hardest parts of the Chemistry examination were considered to be locating the correct chemicals and cleaning up afterwards.

> Cleaning up. Too many people were doing different things so there was no room.

> Supporting all equipment in O2 gas experiment. Not suitable equipment

Students in their other comments on the Chemistry practical examination referred to the perception that the tasks differed in terms of length and difficulty. There were references to how busy examiners were.

> Very good idea

> A change needs to be made if this is to be implemented, be it either the amount of students present for assessment be decreased or more examiners be present.

> Felt examiners didn't observe properly as they were always busy with someone else and not a fair test if someone gets a harder task than another person.

Four students made general comments on the trial.

> I think it is highly necessary to have a practical examination for the Leaving Cert sciences but this examination could be improved.

> I believe that all in all that practical assessment should not be implemented, as it is too difficult to examine and study for properly. Factors such as equipment failing or results being unobtainable, through no fault of the student, tarnish the overall result of the student and cannot be corrected.

Physics

Students found the easiest parts of the Physics practical examination to be gathering the equipment for the tasks and setting it up. There was a wide spread of answers to this question and the next one on the hardest part of the tasks and much overlap.

Found easiest:

> Gathering my equipment and setting up the experiment. Everything was close at hand.

> Writing up experiment. The booklet contained various hints.

Found hardest:

> Carrying out the experiment. The principle of conservation of momentum is hard to carry out alone.

> Remembering what equipment necessary. Had to match the equipment needed in theory to the physical piece of equipment.

> To see the light beam correctly because the room had to be bright for the other students.

In their other comments about the Physics trial, students referred to differences in length and difficulty of tasks, waiting for the examiner, and the good organisation of the trial examination.

> Well organised. However, one examiner monitoring 12 students is impractical, too long, too hard to examine all students at once, etc.

> Some of the experiments were much harder to complete than others. I think the examiners should find a way to even the difficulties.

> It was carried out in a very professional manner.

The general comments were similar to those already made in the other categories. Other comments included the following.

> The practical assessment needs to be in place as soon as possible.

> Everyone doing different experiments makes the examination extremely unfair. The practicals should be done similar to the music practical in which the students can choose which experiments they want to perform.

While there is a general sense of satisfaction with the trial practical examination evident on the part of the students there are some issues that resonate with points made in other parts of this chapter. The first of these is that the examiners were seen to be very busy. The second is that students considered that there were significant differences in the tasks that students were assigned in terms of difficulty and length. They considered also that a small number of tasks in the trial were difficult for one student to do on their own.

4.2 Examiners

In the case of the examiners, detailed information was sought on their experience of examining in the trial and on how well schools were prepared. Looking towards a possible inclusion of practical assessment in the Leaving Certificate examination, examiners were questioned on whether they would be willing to act as examiners and on how practical assessment should be implemented in the Leaving Certificate examination.

Seventeen of the twenty-one Biology examiners, all twelve of the Chemistry examiners, and seven of the eleven Physics examiners returned their questionnaires. The responses of examiners for each subject are analysed below.

4.2.1 School laboratories and designated support teacher

Examiners were requested to comment on the adequacy of the laboratories for the trial and whether they had been adequately set up for the examination. They were also asked about the support provided to them by the designated support teacher.

In general, Biology examiners found that the laboratories were suitable for the trial. Some referred to a lack of space for setting up students' workstations and for the examiner to circulate during the examination. Others commented on equipment that was old and an inadequate supply of chemicals and apparatus.

Almost all the Chemistry examiners stated that the laboratories of the schools had sufficient space. A few suggested that space could be an issue in certain laboratories if

twelve students were to be sitting the practical examination. One examiner said that some laboratories did not have all of the required solutions, chemicals, and basic equipment.

Four of the seven responding examiners for Physics stated that the laboratories did not have adequate space for workstations or equipment but most examiners stated that the set-up of equipment in the laboratories was satisfactory.

Virtually all examiners rated the assistance of the DST in the schools as excellent. All examiners referred to the DST as being essential to ensure orientation in the laboratories. They also referred to the DSTs role in preparing for the practical examination, sourcing the required chemicals and equipment, and replenishing stocks during the examination. DSTs were also reported to have assisted in maintaining student discipline, they ensured the safe working of laboratory equipment and fittings, and they liaised with school management. They also had a major role in managing unforeseen circumstances during the examination.

4.2.2 The trial practical examination

Changing the range of tasks for the examination

Many examiners for Biology, Chemistry and Physics were required to change the range of tasks for the practical examination before the examination took place. This was primarily due to a lack of equipment or chemicals in the case of Biology, equipment in the case of Physics or insufficient coverage of subject content in the case of Biology, Chemistry (in particular organic Chemistry), and Physics. This deficit in course coverage is understandable given the time of the school year at which the trial took place.

The impact of Storm Ophelia on schools was cited by a few Biology and Chemistry examiners as a reason for some schools not being fully prepared for the trial practical examination. One Chemistry examiner noted that it took three people (advising examiner, examiner and DST) two hours to organise the chemicals before the examination could begin. (Because of Storm Ophelia all schools were ordered to close on 16 and 17 October, which had been scheduled as the first days of the trial. For some schools, particularly in the southern part of the country, the closure lasted longer.)

According to one Biology examiner,

> a large number of students had "issues" with the heart dissection task (blood aversion/smell aversion/possible fainting) so these students had to be assigned to alternative tasks.

Supervision of the practical examination session

Students preparing to do tasks

For the first ten minutes of the practical examination session students were required to prepare to do their tasks and gather their equipment. A small number of the Biology examiners reported that there was some overcrowding at the collection points and that circulation around the workstations was an issue. Some examiners reported students showing a lack of knowledge regarding names of pieces of equipment and/or chemicals or having difficulty locating their equipment.

In the case of Chemistry, students' preparations were hindered in some cases through equipment not being labelled or being labelled incorrectly. There were instances of overcrowding at the locations where solutions were stored for the trial, especially in

centres with a larger number of students. There were cases of students taking chemicals that were required by other students or selecting a solution of the incorrect molarity. Other Chemistry examiners noted that where students were familiar and comfortable with the laboratory there were 'no real problems'. Suggested solutions to these problems proposed by examiners included the use of generic containers with a standard labelling system and a designated area in the laboratory for storing and collecting chemicals that contained a limited number of chemicals from which students could choose.

While students were preparing to do the tasks, most Physics examiners encountered students with a lack of knowledge regarding the equipment they needed or having difficulty setting up their equipment.

Students working on tasks

Many Biology examiners stressed that supervision of all students working on tasks while at the same time examining an individual student who was performing a key step of her/his task was the most difficult part of their work. It was suggested that students could be directed to call the attention of the examiner at key steps of task completion and then wait to be observed completing that step.

> Students doing task too quickly and having steps performed before examiner getting opportunity to assess the step. This phase caused most difficulty. Trying to monitor and assess students.

The issues for the Chemistry examiners in supervising this phase of the practical examination were similar. Getting around to all students was the main difficulty, especially where there were 12 students in the laboratory. Monitoring key moments of the different tasks was an issue for one examiner. Solutions suggested included marking the steps in each of the task booklets where the examiner wanted the students to pause their work and call the examiner's attention. Another examiner suggested that the students could be asked to explain the steps that had not been observed by the examiner. Some examiners stated that 8–10 students should be the maximum number permitted to sit the practical examination at one time.

> In chemistry, almost 80% of the time was spent carrying out the practical, which meant the exam was focused almost entirely on practical skills being observed by the examiner. This was good. Care should be taken to ensure equal distribution of practical content as some tasks took longer than others. Examiners found it difficult to observe key moments in the task and this required verbal interaction between the examiner and student to agree to wait for examiner to be available to see these moments. These times should be highlighted in the task booklets where necessary.

The main difficulties that arose regarding supervision of this phase for Physics examiners were related to observation of students' performances of key steps of the experiments. There were some safety issues (e.g. heat experiment) referred to and situations where students did not select the appropriate equipment. There was some consensus among examiners that students should be directed to call the attention of the examiner at key steps of task completion and wait to be observed completing that step. Some examiners found that being fair to all students and judging the marking process (e.g. when to apply four marks penalty or observe accuracy of student measurements) was difficult.

> Difficult to see the exact measurements being taken. Would be more helpful to ask the student to explain their work. Students also refrained from calling you at the key moments during the task, even though it was highlighted in the task booklet. Some students struggled with which measurements to take, it was hard to judge the level of detail you should give them for the (-4) marks [for help given by the examiner]

Students writing answers and finishing up and clearing their workstations

Biology examiners reported no particular issues or difficulties with supervising students completing the written section of the examination or finishing up and tidying their workstations.

In the case of the Chemistry practical examination, some examiners reported a lack of adequate time for students to complete the written part of the examination. This was because some students commenced this part last and some tasks were quite long. Examiners suggested solutions to reduce the time taken by students in the practical examination. These included pre-testing of chemicals, e.g. ensuring that the hydrogen peroxide was fresh; including pre-drawn graph axes in the task booklet; and including just one single task. Most examiners did not experience any difficulty with students finishing up and clearing their workstations.

Most Physics examiners reported that this phase was quite straightforward to supervise. Among the issues that arose were students not having data to write up and students not knowing what formula to use. There were no issues or difficulties with supervising students finishing up and tidying other than in one case where the equipment was too hot to put away.

4.2.3 Assessment of students' practical skills by direct observation

Students were observed by the examiner as they carried out their tasks and awarded marks in accordance with a marking scheme. The marking scheme that was used by examiners was a generic one, as outlined in Chapter 3. The same marking sheet was used for each subject and task by examiners to record their marks for students.

The assessment of students' practical skills by direct observation while they carry out their tasks is the core of the trial. Examiners were asked to give their judgements in three areas: whether examiners used methods of recording students' marks in addition to the form provided; how they made judgements on students' performances and awarded marks; and whether they had suggestions as to how this directly observed part of the assessment could be improved. These areas are addressed separately for each of the three subjects in the following sections, and the material is based on examiners' responses.

Biology

Most examiners used additional methods or materials to support their implementation of the marking sheet. These methods included recording notes in the margins of the sheet or on an extra sheet and using their own shorthand codes. Some examiners used a grid note-taking system to track students' performance. Examiners then combined their grading systems and any other notes to inform their overall mark once the tasks were completed.

The following are some examples.

> Recorded notes at side of marking sheet. Highlighted important points for each task I needed to see each student do.

> A sheet with boxes for each experiment with space to write comments. In this, I could write notes and use these notes to decide on my score after the session was over.

> Developed a shorthand for each section and each practical to remind myself of key things to look for when I was getting to grips with how the examining was working.

The main issues for the Biology examiners in judging students' practical performance were monitoring their progress at appropriate intervals and the use of the grading system. Some examiners said that they had missed the moments when certain students were performing key steps. Suggestions to address this, as referred to earlier in the chapter, included asking students to call the examiner's attention when they reached critical stages. Some examiners found that it was difficult to make judgements in relation to minor and major errors made by students while endeavouring to give credit for the overall performance of the students. One examiner stated that it was important to reserve judgement of a student's performance until the practical examination was completed because a student might correct their errors at a later stage in the examination.

As a means of addressing the difficulties in direct assessment, examiners suggested more detailed assessment criteria, taking more time to make sure everything was ready before the examination, and instructing students to call the examiners at key stages in their task. Other examiners said that they would prepare a more detailed marking sheet that would include 'clear points to look for in each task noted down on a grid beforehand'. Examiners referred to the importance of training in assessment by direct observation.

In the training of examiners for the trial examination, they were requested not to speak with students unless it was necessary. In the event, some examiners did find it necessary. In responses to the questionnaire, examiners said that they should be allowed to speak to students during the exam to find out what steps they had completed in their tasks. They also said that students should be instructed before the examination to make sure that the examiner is present when they are conducting key steps.

Chemistry

As with the Biology examiners, almost all Chemistry examiners used short-hand codes and a checklist system to supplement the marking sheet for grading students' work. In the case of the Chemistry examiners, the checklist included checking each reagent used, each step in the experiment, and key observable moments of the experiment. Examiners also had a sheet of paper with spaces for each student and comments for each student or they checked students' performances against a list of key skills and observations.

Some examples from the Chemistry examiners

> In the week prior to the exam, I read through each of the tasks and listed all reagents required and steps to be completed in order. As the students completed their tasks I ticked each correct reagent placed on bench and each step observed to be taken and placed an X and sometimes a note besides any step missed or

> incorrect....The students were graded at the end of the session and these recordings were important for fairness and accuracy.

> Advising examiner had supplied a list of key skill/observations for each task. This was attached to a clipboard on the left-hand side. When a skill was observed, it was ticked on the list. If it was omitted or incorrectly performed I placed an x beside it.

The Chemistry examiners considered that making sure that there was accuracy and fairness in the grading process for each student was the main challenge in judging students' practical performance. For many the supplementary resources they put together helped to ensure that accuracy and fairness.

When asked what they would do differently if they were to repeat the process, the following were given:

- check and arrange all equipment and chemicals for ease of access and use on the day of the examinations
- ask students to get the examiners' attention when they reached certain stages in the experiment
- make use of a comprehensive supplementary marking sheet [as referred to in previous paragraphs]
- have fewer students in the laboratory
- ask for an image of the laboratory in advance of the examination day.

Examiners said that they found examining in the trial to be a 'very positive experience'. There were frequent references in questionnaire responses to twelve students being too many in a centre. One examiner suggested that as part of their preparation, examiners should carry out the practical tasks and agree the standard and requirements expected of students.

Physics

Most examiners reported utilising additional methods or materials to support their implementation of the marking sheet. These methods included recorded notes in the margins of the sheet or on an extra sheet and employing short hand codes such as circles, dots and ticks. Two examiners did not use extra sheets.

Some examples from the Physics examiners

> Needed space to write things as I was seeing them. Had a box for each heading and made notes on things as I saw them that gave me the opportunity to mark later on which took into account the students righting wrongs as they went along.

> System of dots for omissions or inaccuracies, ticks for positive steps taken. Mark awarded at end of appropriate stage. Fatal flaws indicated by a circled -4. Mark awarded based on the series of dots and ticks for each assessment item at end of appropriate stage.

The primary difficulties for Physics examiners as with the other examiners were in relation to the grading process and the observation process. They suggested that it was difficult to translate the extent of assistance received by some students into specific negative marks in

grading. Other examiners related that they found it difficult to observe the accuracy of students' measurements during the practical.

> I found it difficult to get around to see students when making measurements. In these cases, I asked students to repeat what they had done.

> Sometimes difficult to differentiate between the different headings for marking. Assembly and use for example or use and measurements. With some measurements, it was hard to tell if they were measuring the right thing without asking the students to describe their method. I think that asking them to describe what they did is a good way of aiding the marking.

Some examiners noted the difficulty of observing all students' performance simultaneously and stressed the importance of students pausing at critical stages in the experiment for the examiner to be present at their workstation to observe critical procedures being conducted. A few examiners pointed out that students working with equipment could potentially have issues such as breakages and having to deal with faulty equipment.

4.2.4 Impact of practical assessment on teaching and learning

Examiners were asked their opinion on the impact they would expect practical assessment to have on how teachers teach. Most examiners thought that the practical assessment would result in more student practical work in classes and a greater emphasis on individual student practical work. Many examiners felt that practical assessment would be included in house exams in schools.

There was some consensus among examiners that having a practical, hands-on, and more realistic experience of science would engage students more and encourage them to take ownership of their practical work. Examiners also looked at other possible effects of having increased student practical work in schools. These included trying to facilitate and manage individual student practical work in large classes, greater pressure on laboratory access and more potential health and safety issues. A few examiners noted that current laboratory equipment might be insufficient and one noted that 'the call for lab technicians would grow louder'.

4.2.5 Impact of practical assessment on students' attitudes and engagement

Examiners gave their opinions on how having a practical examination in the Leaving Certificate would affect students' attitude to studying Biology, Chemistry and Physics.

Examiners considered that practical assessment would enhance students' engagement with and attitudes towards conducting practical work. It would also develop their skills in manipulating equipment, using specific apparatus and making up solutions. One examiner stated that

> It would be a huge advantage [for students] to understand how the scientific method, good lab practice and independent learning are vital should they continue their studies in Science.

They said that practical assessment would result in students being encouraged to engage with the subjects and to improve their 'deductive, experimental and analytic skills'. They

considered that it would promote their confidence in the subjects and encourage them to take responsibility for their learning. It would also develop their problem-solving skills.

> I think from chatting to my own students they would welcome an introduction to practical assessment. It would inspire them to take more responsibility for their own involvement in practical class.

4.2.6 Fairness of the tasks and the examination

Overall, most Biology examiners thought that the practical examination was fair. Approximately half of the examiners considered the tasks themselves fair. However, most examiners felt that there were discrepancies in the level of complexity, the challenge and the duration of the various tasks. Some examiners suggested the equity of the examination would be enhanced if clearer instructions were provided to examiners regarding the marking process in the direct assessment.

> It could be fairer if examiners were given clear instructions on what to look out for in each experiment and what to deduct marks for.

> Level of difficulty and duration of each need to be measured carefully to ensure the tasks were fair.

> Allow flexibility ... by assessing general practical skills, e.g. observing meniscus etc. than actually prescribing specific tasks. General task skill assessment may also have the benefit of reducing risk of rote learning of the tasks, which may be a concern if the current model was rolled out.

Generally, Chemistry examiners felt that the practical examination was fair. Most examiners felt that the tasks were fair. Some examiners felt that perhaps some tasks required a longer time to complete. A few examiners felt the need for a more detailed marking scheme for the direct assessment of students' performance.

> Indirect assessment--very fair. Direct assessment: a tighter more detailed marking scheme is required possibly as outlined above, a list of key skills that should be observed and how many they should perform in order to gain full marks; how penalties should be applied. If they perform almost all skills correctly but have made one or two errors or omissions do they still get full marks?

Many Physics examiners thought that there were issues with the perceived fairness of the tasks in the examination. They considered that some tasks were more difficult than were others. Some examiners said having the same examination for Higher Level and Ordinary Level students could be unfair to Ordinary Level students.

> I think the tasks themselves were fair. However, I thought the marking scheme made it quite difficult for the ordinary level candidate who was expected to answer at a much higher level than what would be expected in written exam.

> There was also a big focus on precautions, with some tasks asking for more than three precautions or sources of error. That is very tough for a lot of students.

> I feel some tasks were more difficult than others (e.g. heat) maybe some of the more difficult tasks could be more guided in the booklet in terms of set-up or measurements to be taken.

> The tasks were fair but they were not equal. Some tasks could be extended somewhat e.g. Boyle's Law.

4.2.7 Practical assessment in operation

Examiners were questioned on the practicalities of a national implementation of practical assessment, including the availability of examiners, and the best time of the academic year for the examination.

Examiners for practical assessment

Most (92%) respondent examiners stated that they would apply to be an examiner for a Leaving Certificate practical examination. Lack of suitably qualified substitute teachers to cover classes was seen as a drawback. One examiner stated that they did not think the school principal would permit such an absence every year. Another examiner gave as a reason for not applying that they 'were not confident that it will be possible for the process to be fair' and that they would 'prefer to be in school with my students when the process is ongoing'. Another examiner said he/she would not apply to become an examiner because of the quantity of paperwork involved versus the remuneration.

Is 90 minutes with 12 students and 3 sessions a day appropriate for the practical examination?

Most examiners considered 90 minutes to be an appropriate length for the practical. All examiners would change the maximum of 12 students in the practical by decreasing the number of students undertaking the practical simultaneously. The average number suggested was eight students. Taking the three subjects together, most examiners said that it was feasible to hold three practical examination sessions in one day, maybe not on the first day in a school.

Minimum length of time required for the interval between practical examination sessions

Biology examiners suggested an inter-session break time ranged from 20 minutes to 90 minutes. A one-hour break would suit most of the Biology examiners. Chemistry examiners would require a longer break between sessions with their suggestions ranging from 45 minutes to two hours. A break of 90 minutes would suit most of the Chemistry examiners. In the case of the Physics examiners suggestions ranged from 30 minutes to three hours with a break time of one hour suiting most.

Proportion of marks for the practical examination

Most examiners considered the 30% weighting to be appropriate with the primary reason being to give recognition to the importance of students' laboratory skills and to give credit for the class time involved in student practical work. Of the two examiners who considered the 30% weighting inappropriate, an alternative of 15% was offered.

Most examiners considered the 50:50 division of marks between direct assessment and indirect assessment to be appropriate and the main reason given was the importance for students of understanding the theoretical basis for the practical work being undertaken.

> As the assessment is common, it is the most fair to all students. It allows for differentiation of questions in the indirect component to adequately assess Higher and Ordinary Level students.

Timing of the practical examination

It is interesting that while some examiners suggested other times in the academic year the majority of examiners (66% of those who responded) considered that the practical examination should take place between February midterm and after the Easter holidays.

Reflections on their role as examiners in the trial

Most examiners considered the most rewarding part of their role to be the interaction with other schools and professionals and observing students as they participated in and enjoyed practical assessment. Many examiners reported that they benefited from being in different schools and from learning about how those schools conducted their daily business and managed their laboratories. Another important aspect of the role, as described by some examiners, was interacting with students and observing them enjoy the experience, gain a sense of achievement, rise to the challenge and develop their confidence.

> Excellent CPD, great to meet and learn from like-minded individuals who are committed to improvements and modernisation of the LC curriculum.

> Seeing the students confidently gather equipment and getting on with the task at hand…this could be part of something big!

> Seeing wide variety of approaches, engaging with teachers and students. Broadening of my own outlook on the strategies and methodologies that may be used in teaching the subject. I recommend such work be classified as CPD or all participants and the experience rotated among as many teachers as possible

The most difficult part of their work for most examiners were assessing students' work and becoming familiar with the equipment in the school. Some examiners found time management and liaising with schools to be a challenge. Examiners noted that endeavouring to be fair to all students was difficult while marking them based on observing as much as possible of their work, as up to twelve students were carrying out tasks simultaneously. Examiners stated that they felt a lack of knowledge in answering all of the questions they were asked, and found orientation in each new laboratory a challenge.

> Finding accommodation in Dublin with such short notice between the Oct 7th conference and examining (not helped by Ophelia)

> Trying to carry out the assessment without making the students feel uncomfortable and remembering each of the key elements of each task that I needed to look out

> Steep learning curve for me, but I enjoyed it all. Quite difficult trying to gauge what was worth marks on the written paper. I think because this paper was common level, it was hard to find a balance between the HL and OL scheme.

Generally, examiners were positive about the potential impact of this practical examination on students' learning.

> Would be such a great thing if practicals became part of LC Chemistry. Would benefit all students of all abilities, attract more students to Chemistry and give credit for all the practicals that are done in Chemistry

> Delighted to be involved…learned a lot that will benefit my own teaching…I hope that practical examinations are introduced as Science is a practical subject. All

> other practical subjects have a practical component. It is desperate to see the JC reducing the practical element.

4.3 Centre reports

Examiners and advising examiners were requested to completed centre reports to provide information on the experience of the examiner in the school during the trial. The material in this part of the report is based on the forms returned by 17 Chemistry examiners and 10 advising examiners.

Contact with DST for your subject

All examiners conveyed that they had sufficient contact with the DST regarding the examination process. The nature of this contact primarily included phone calls to discuss and organise the examination sessions, meeting with the DST before and following the examination, and contact with the DST during the examination when required. Some examiners reported that contact was made with the DST during the exam either by way of text message, phone calls or by calling upon the DST because he/she was in close proximity to the laboratory in which the examination was taking place (e.g. in the prep room, or in the corridor).

> I first made contact with the DST at the end of the marking conference. We were in touch by phone and text. Spoke after each session in prep for next session and she was available all day to top up chemicals and deal with other situations which may arise. She could not be more helpful.

In most cases, the DST was a teacher of the subject being examined. The DST introduced the examiner to the school's laboratory for the examination. Prior to the examination, each examiner had a briefing meeting with the students in a designated area such as a demonstration room, the laboratory, a classroom or the school canteen. The agenda for the briefing meeting was included in the detailed instructions that were given to examiners. The purpose of the meeting was to introduce the examiner and to explain the examination process to students. The usual length of the meeting was 10-20 minutes. In general, students asked few questions at these sessions. In most schools teachers attended.

> Fabulous facilities; Chemicals stored in fume hood …and on a trolley which was a bit crowded; Equipment and glassware stored in presses; No helpful material on wall; Gas and electric cut off, safety kit;

The practical examination

In general, there was minimal interaction between students in a session; they shared some pieces of equipment e.g. electronic balances cooperatively and non-verbally and otherwise worked completely independently. The primary issues with students collecting their equipment for the examination were due to overcrowding at the collection point (trolley, bench, etc.). Examiners noted that students seeking help from the examiner diverted their attention resulting in the examiner missing key moments of work of other students. The point was also made that it was difficult not to engage with students during the examination. Examiners had to resist the tendency as teachers to point out errors and correct students in the trial when they observed them making mistakes and to adhere strictly to their role as external examiner. Examiners had to minimise interaction when students spoke to them. Examples of such engagement were locating chemicals for

students, supplying graph paper as rough work paper, and helping them to light the Bunsen burner.

> No interaction between students during trial; In first session with 12 in the room I found it chaotic too many people around the chemical bench at the same time; They weren't able to locate the chemical needed due to some students taking it to their own benches. I found the chemicals for them; The session was difficult to monitor; The experimental aspect was straightforward; The part I found difficult in the session is to get around to the different tasks at the right times in order to monitor them. Observing the washing of the apparatus for dilution or volumetric analysis was difficult to monitor as these can be done quickly. Students were asked to repeat these parts if they had not been seen.

For examiners who had a subsequent session on the same day the issues that arose between sessions included the pressurised nature of the interval between sessions due to cleaning up and restocking chemicals.

Most examiners felt that either it would not be possible or it would be challenging to conduct three sessions in one school day due to the considerable preparation time and the level of concentration required. Most examiners agreed eight or nine students should be the maximum number in a session, especially if all students were doing different experiments.

> If students were familiar with the layout and equipment storage arrangements of the lab, 3 sessions feasible. Two session on day one would be feasible as time is required to check chemicals/equipment etc. in the morning of day 1. This took between 1.5 to 2 hours.

> Three sessions is too much. I found it very challenging.

Training of examiners

Most examiners described the training they received as very good or sufficient and the documentation as easy to follow.

> The instructions need to be simplified and refined but it was great to have documentation to refer to.

> Use same paper codes for hard copy and electronic versions.

Chapter 5 Evaluating the technical quality of the assessment – quantitative evidence

The trial produced a comparatively rich data set that facilitates exploration of a number of issues related to the potential validity, reliability, discriminatory power, and fairness of the assessment. There are many statistical procedures that are used by assessment specialists to analyse certain aspects of the quality of tests. This chapter reports on the outcomes of applying some of these techniques to the data gathered in the trial. This can help us decide whether it is a good test from this technical perspective.

5.1 Marking reliability

One issue of concern when considering any new type of assessment is the degree to which it can be reliably marked. Candidates' scores should, to the maximum extent possible, reflect only their performance on the day, and not be influenced by extraneous factors, such as the leniency, severity, or inconsistency of a particular examiner.

Marking reliability is the degree to which different examiners will independently award the same mark to the same performance. The marks awarded by direct observation on the day and the marks awarded for the work in the booklets were two distinct operations. The booklet work was marked in the same way as written papers are, so the question of reliability in the marking of these, along with the procedures required to ensure adequate levels of reliability, form well-trodden ground. That is, we already know that this can be done sufficiently well. Accordingly, of most interest here is the degree to which the marking in real time of directly observed performance can be carried out reliably.

One comparatively direct way to measure marking reliability is to arrange for a selection of performances to be independently double-marked. This approach was used in the trial. More advising examiners than would usually be the case were appointed. The particular form of monitoring that was implemented involved the advising examiner being present for entire sessions and independently marking the performance of the candidates. Neither the examiner nor the advising examiner was aware of the marks that the other was awarding[4]. Although it may be the case that the advising examiners were more skilled in their assessments than the examiners, which means that this model of double-marking is not exactly the same as having two randomly selected examiners, it is a sufficiently close approximation for our purposes.

The correlation between the pairs of scores for all double-marked performances is a measure of marking reliability. A correlation of 1 indicates perfect agreement between all examiners concerned, and a correlation of 0 indicates no association[5] whatsoever between the marks awarded by the different examiners, suggesting complete randomness in marking.

[4]Advising examiners saw the marks that an examiner had given only after completion of the session and the marking. This information may subsequently have been used to moderate the examiner's marks.
[5] Strictly, no linear association.

The correlations found in the trial for the directly assessed observations of practical performance were as shown in table 5.1 below.

Table 5.1 Marking reliability for assessment of practical skills by direct observation

Subject	Number of double-marked performances	Correlation
Biology	211	0.806
Chemistry	113	0.875
Physics	93	0.921

These are high levels of correlation for an assessment of this type. This indicates that examiners working in this context can indeed achieve a level of marking reliability that is adequate for the purpose.

5.2 Discrimination

Discrimination refers to the degree to which a test instrument, or any individual element of it, is effective in separating out candidates with differing underlying levels of achievement. Irrespective of whether average marks are high or low overall, if the candidate marks are not adequately spread out, the test is not helping us to distinguish strongly between higher- and lower-achieving candidates. In an assessment of the form under consideration here, having a reasonable spread of marks is a necessary but not sufficient condition for adequate discrimination. The spread of marks could be caused by random effects or by attributes that are irrelevant to the one we are measuring. If we can be confident that the various elements of the test are focused properly on the attribute we are trying to measure, then inferring adequate discrimination from an adequate spread of marks becomes more reasonable. Mark distributions are relevant to evaluating the discriminatory power of the test, and so are treated here.

The distributions of marks awarded in each of the two parts of the assessment (*practical skills* and *results and analysis*), as well as that of the overall marks, is summarised for the three subjects in figures 5.1 to 5.3 and tables 5.2 to 5.4 below. While marks were awarded out of 60 for each part, yielding a mark out of 120 overall, marks are expressed as percentages in these figures and tables.

In all three subjects, it may be noted that the marks are much higher for the assessment of practical skills than for the assessment of results and analysis, with high numbers achieving in excess of 90% of the available marks. Indeed, the strength of the distribution for the practical skills results in considerable skewness[6] and a ceiling effect[7]. This suggest that this aspect of the assessment is not as discriminating as it might be, particularly as regards its capacity to discriminate among stronger candidates. On the other hand, in a criterion-referenced or standards-based assessment system, as distinct from a norm-referenced one,

[6] The skewness of a distribution is, in broad terms, the degree to which it is asymmetrical. Test score distributions usually have many scores somewhere near the middle and fewer on either side. A skewed distribution has many scores bunched to one side, with a tail extending to the other.

[7] A 'ceiling effect' is when significant numbers of students achieve full or almost full marks in a test. The test then cannot discriminate between these candidates who have all reached or exceeded the highest level of competence that the test can measure.

it can be argued that this is not necessarily problematic. It could be argued that if we are satisfied that a large majority of students have reached or exceeded the standards expected of them in a particular subdomain, then a low level of discrimination among those students is not necessarily problematic. Nevertheless, the question of whether the level of practical skill expected of students was sufficiently challenging warrants consideration.

Notwithstanding the strength of the mark distribution for the practical skills part, the overall mark distribution, while strong, does not appear problematically so.

Figure 5.1 Mark distributions for parts and total of the assessment - **Biology**

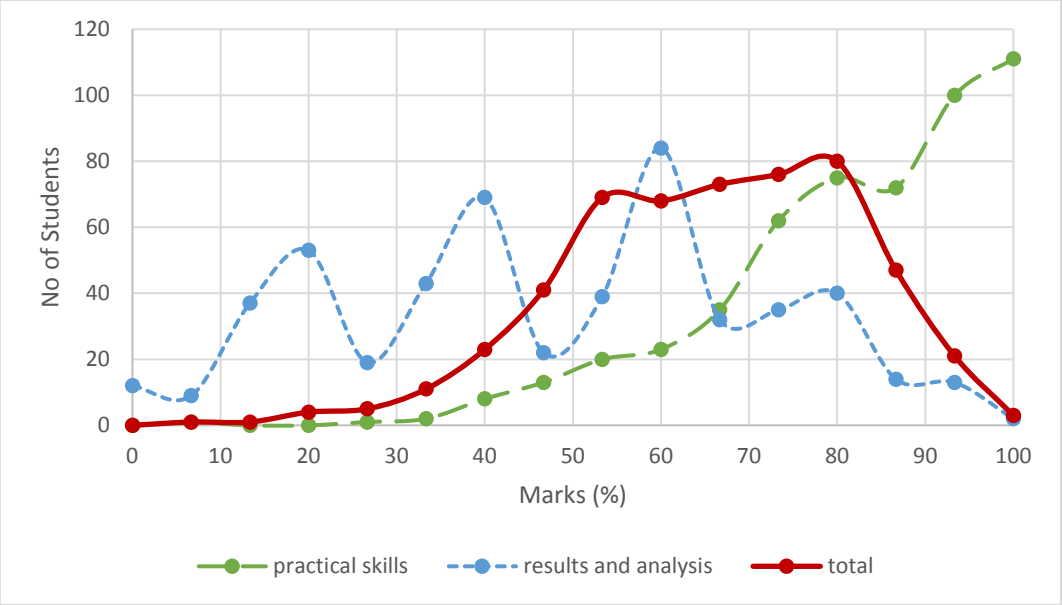


Figure 5.2 Mark distributions for parts and total of the assessment - **Chemistry**

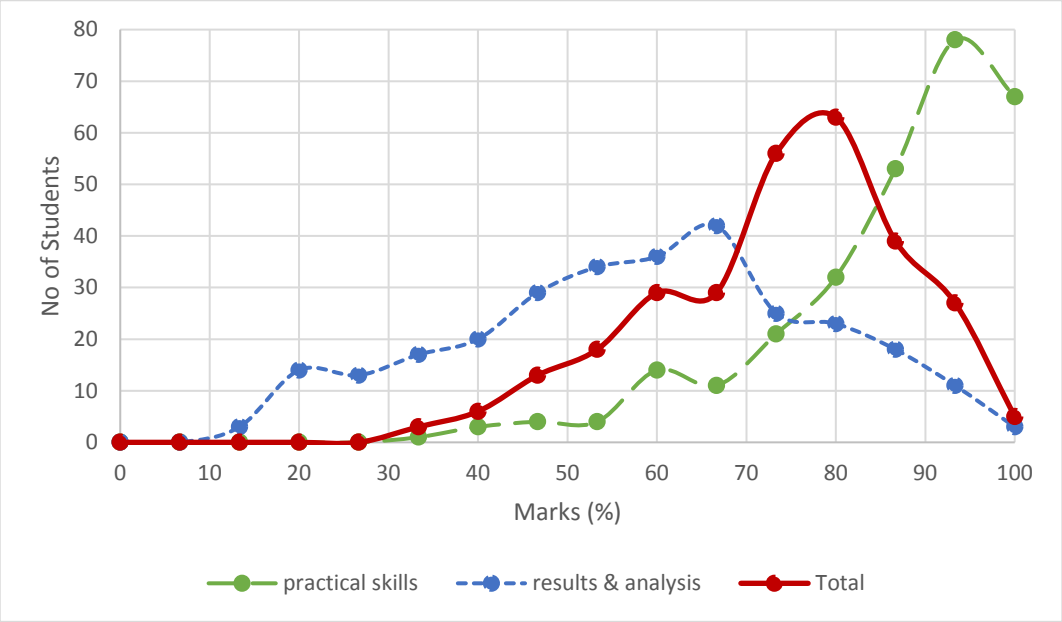


Figure 5.3 Mark distributions for parts and total of the assessment - **Physics**

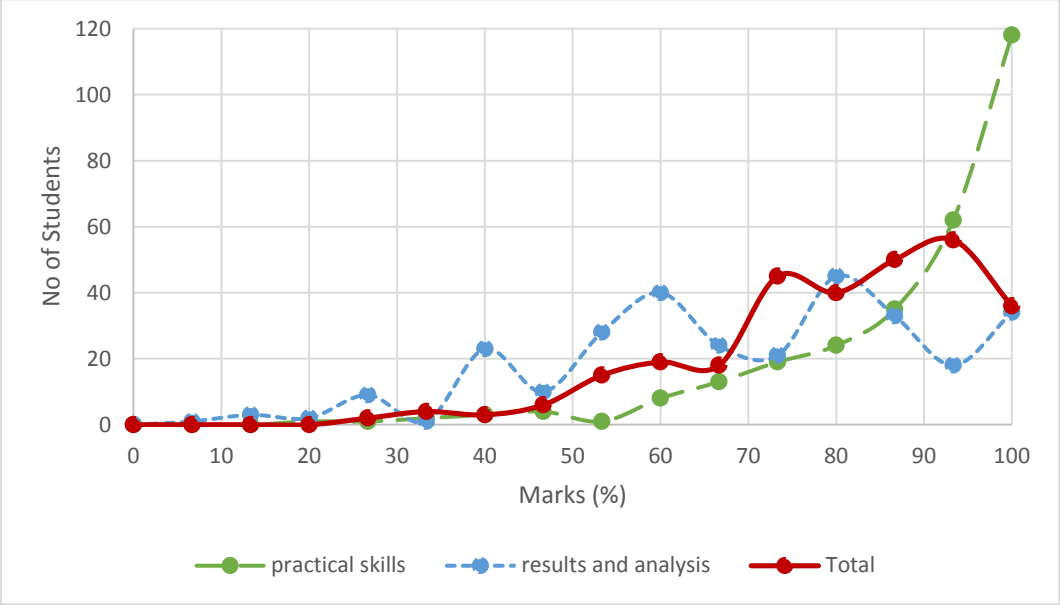


Table 5.2 Mark distributions for parts and total of the assessment - **Biology**

	Average mark (%)	Range of marks (%)	Standard deviation (%)	Percentage of students scoring above 90%
Practical skills	80.2	6.7 - 100	16.2	37.0
Results and analysis	45.7	0 - 95	23.3	2.7
Total	63	3.3 – 98.3	16.2	2.3

Table 5.3 Mark distributions for parts and total of the assessment - **Chemistry**

	Average mark (%)	Range of marks (%)	Standard deviation (%)	Percentage of students scoring above 90%
Practical skills	85.2	31.7 - 100	13.9	50.3
Results and analysis	55.7	10.0 – 98.3	20.0	3.8
Total	70.4	30.0 – 98.3	14.2	5.6

Table 5.4 Mark distributions for parts and total of the assessment - **Physics**

	Average mark (%)	Range of marks (%)	Standard deviation (%)	Percentage of students scoring above 90%
Practical skills	87.9	20 – 100	15.3	59.5
Results and analysis	66.8	5 – 100	20.9	17.0
Total	77.3	25 – 100	15.4	26.3

5.3 Achieved weights

When an overall examination consists of different components that are intended to measure distinct aspects of the target domain (such as oral language skills as distinct from written language skills), a decision is made as to how these distinct sets of skills ought to be valued relative to each other in combining the results of the components into an overall result. The allocation of marks to the different components is typically taken to represent a set of 'intended weights' – a numerical expression of the intended relative values to be assigned to the components.

However, when examination components are actually implemented, the relative impact of the component scores on the students' overall scores is not necessarily in proportion to the mark allocations. The relative impact of each component on the overall score is very much dependent on the spread of marks within the component. There is considerable literature on the subject, and some have argued that, because of the influence of mark spread, the standard deviations of the components should be taken as the measure of their achieved weights. However, this is problematic, and the most commonly used measure of achieved weight is one proposed by Adams and Murphy (1982)[8].[9]

Two issues related to achieved weighting are of interest here: the weightings of the two parts of the practical assessment relative to each other, and the likely relative weights of the practical and written components in the event that the practical assessment is rolled out as part of the Leaving Certificate examination. We can say a lot about the first of these and something at least about the second.

5.3.1 Achieved weights of the two parts of the practical assessment

The achieved weights (relative to each other) of the two parts of the assessment for each of the three the subjects are shown in Table 5.5 below, the weights having been calculated using the Adams & Murphy formula.

Table 5.5 Achieved weight of each part of the practical assessment

Subject	Achieved weight of *practical skills* part	Achieved weight of *results and analysis* part
Biology	0.37	0.63
Chemistry	0.37	0.63
Physics	0.42	0.58

If it can be assumed that the intention in allocating marks equally between the two parts was to make each of them have a similar impact on the overall score, then the achieved weights can be interpreted to mean that the practical skills element has been underweighted relative to its intended value. Nevertheless, caution needs to be exercised in such an interpretation. The achieved weighting for the practical skills element has clearly been heavily influenced by the high scoring in this component, which has compacted the

[8] The achieved weight of component *A* is the standard deviation of component *A* multiplied by the correlation between component *A* and the total score, divided by the standard deviation of the total score.
[9] For a fuller discussion of intended and achieved weights in the context of the Leaving Certificate examination, see Millar et al (2006).

distribution into the upper end of the mark range, reducing its spread, as discussed in 5.2 above in the context of discrimination. As with discrimination, some might argue that achieved weighting is a concept associated with a norm-referenced paradigm – the measure is fundamentally based on how candidates are performing relative to each other within each part and is independent of any external objective standard. Such an argument would suggest that if the score distribution for the practical skills component is defensible on the grounds of intended standards, its impact on the achieved weight should simply be accepted. For the sake of clarity, however, it would be useful to establish whether the equal mark allocations were explicitly intended to represent intended weights in the sense captured by the Adams and Murphy formula, or whether they are simply mark allocations to be taken at face value and without such implication.

5.3.2 Likely achieved weight of the practical and written component

Since we do not have Leaving Certificate written examination results for the students participating in the trial, we cannot measure the achieved weight of the practical component relative to the written one. However, as has been noted, achieved weight is heavily influenced by standard deviation of the scores. If the practical assessment and the written examination have similar levels of spread in percentage terms, then they are likely to combine with achieved weights that are close to their relative mark allocations, although we cannot be sure.

Table 5.6 Spread of marks in practical assessment (trial) and written examination (Leaving Certificate)

Subject	Standard deviation of practical assessment total mark, expressed as a percentage	Standard deviation of 2017 Leaving Certificate Higher level scores, expressed as a percentage
Biology	16.2	22.1
Chemistry	14.2	21.4
Physics	15.4	18.5

In the case of Physics, it seems likely that the achieved weight of the practical assessment component will be a little below the nominal 30%, and somewhat further below for Biology and Chemistry.

5.4 Convergent and discriminant evidence of construct validity

Broader considerations related to the validity of the assessment are dealt with elsewhere in this report. Nonetheless, it is appropriate to consider the extent to which the data gathered can provide any quantitative evidence in support of or against the construct validity argument[10].

Two forms of evidence can be considered in the context of the data gathered – convergent and discriminant evidence. If a new form of assessment is intended to measure a particular

[10] *Construct validity* is the degree to which the test procedure is measuring the construct or attribute that we want to it measure. The argument-based approach to validation provides a methodology for establishing the validity of a test through a process of making explicit the intended interpretation and uses of the test scores and proceeding to gather, present, and interrogate the evidence that supports such interpretation and use.

attribute, we would like it to agree with other measures of the same attribute and not to agree too strongly with measures of attributes that are known to be different from the one we intend to measure. The first of these (agreement with other measures of the same thing) is referred to as *convergent validity evidence* and the second (disagreement with measures of a different thing) is *discriminant validity evidence*.

There is limited potential for convergent evidence in the current context. Alternative reliable measures of the students' practical skills are not available. Nevertheless, for the trial, the students' teachers were asked to provide two pieces of information in respect of each student: a percentage mark representing the student's overall result in their written examination in the subject at the end of the previous school year ('summer test') and, distinct from this, an assessment on a three-point scale (high, moderate or low) of each student's level of skill in practical laboratory work. If the teachers' assessment of their own students in each of these two respects is accurate, and if practical science skill is really a distinct attribute from overall achievement in the subject, then we would hope to see the results of a practical test agreeing strongly with the teacher assessment of practical skills and not so strongly with written test performance in the subject. Such observations would yield (respectively) convergent and discriminant evidence of test validity.

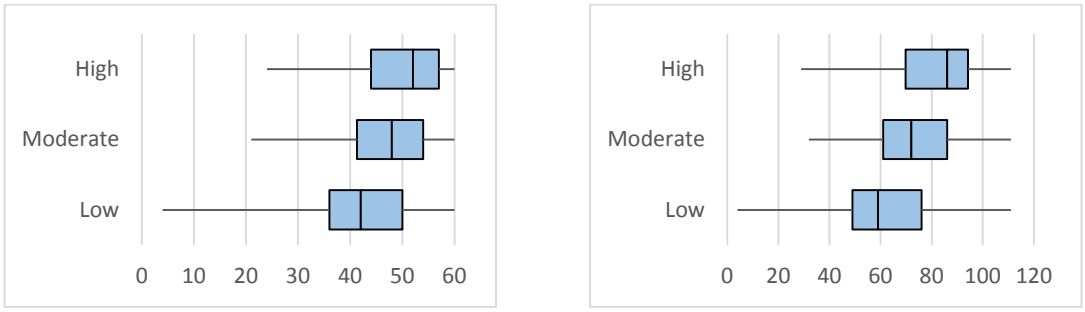
While this is intuitively a straightforward idea, its applicability is limited by the quality of the alternative measures to which we are comparing. This is expanded on below, but first it may be noted that the features one might hope for, as described in the last paragraph, were not strongly observed.

In the case of Chemistry and Physics, the teachers' prior assessment of practical skills (high, moderate, or low) displayed only a weak association with the subsequent performance of the students in the directly assessed element of the test, and a slightly stronger association with the total mark (practical skills + booklet). This is illustrated in the boxplots[11] in Figure 5.5 below (Chemistry) and Figure 5.6 below (Physics). In the case of Biology, the teachers' prior assessment of practical skills displayed a stronger association with the subsequent performance of the students in the directly assessed element of the test than in Chemistry and Physics, and again the level of association it displayed with the total mark was stronger. This is illustrated in the boxplots in Figure 5.4 below.

For Chemistry and Physics, the summer test mark was actually a stronger predictor of both the directly observed skills score and of the total score than was the teacher's assessment of practical skills[12]. This was not the case for Biology, where the teacher's assessment of practical skills was the better predictor of the directly assessed element and was about equal to the summer test in predicting the total score.

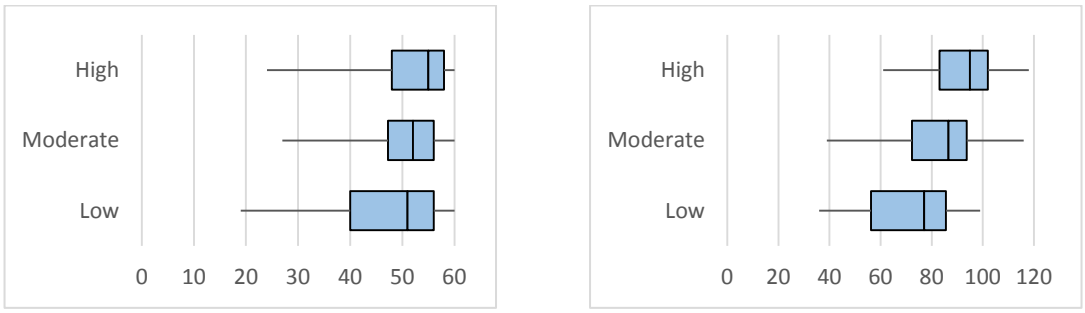
[11] A boxplot (or box and whisker plot) illustrates a distribution by presenting five key (non-parametric) statistics: the median, the quartiles, and the maximum and minimum. The shaded rectangle goes from the lower to the upper quartile, so that it represents the "middle half" of the data. The line dividing the rectangle in two is the median, and the "whiskers" extend from the quartiles to the maximum and minimum values in the data set.
[12] For fair comparison, these "summer marks" were converted to the same High/Moderate/Low scale as the assessments of practical skills, placing a similar number into each category as there were in the practical skills assessment data.

Figure 5.4: Distribution of scores for students whose practical skills were rated in advance by their teachers as high, moderate, or low – **Biology**



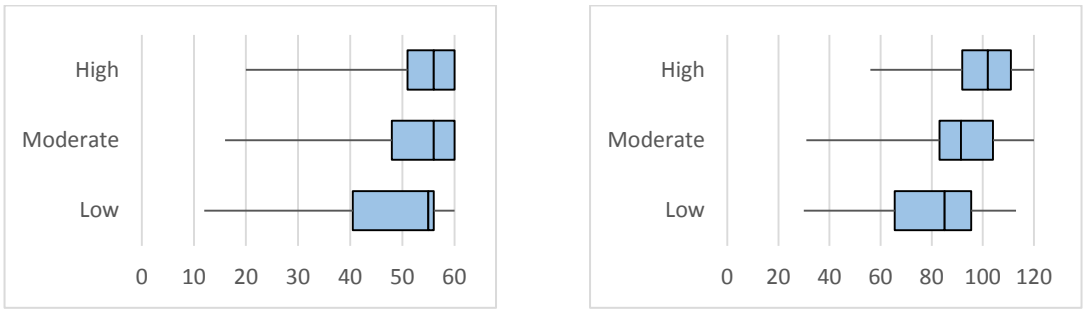
Directly observed mark *Total mark*

Figure 5.5: Distribution of scores for students whose practical skills were rated in advance by their teachers as high, moderate, or low – **Chemistry**



Directly observed mark *Total mark*

Figure 5.6: Distribution of scores for students whose practical skills were rated in advance by their teachers as high, moderate, or low – **Physics**



Directly observed mark *Total mark*

Accordingly, the analysis yields comparatively little convergent or discriminant evidence in support of the construct validity of the assessment. Nevertheless, absence of evidence is not necessarily evidence of absence. In this case, there are a number of very plausible reasons why we might observe lower correlation than we might like between this practical test and the teacher assessment of practical skills, and higher correlation between the practical test and the summer test. For two measures of an attribute to correlate strongly, each must have high reliability in its own right. If one of them is very reliable (and valid) and the other very unreliable, they will not correlate strongly. If we had observed a high correlation between the teacher assessments and the trial test scores, this would have

provided strong evidence that they were both good measures of the same thing. But observing a low correlation tells us only that one or other is poor (or both are).

It is plausible to suggest that the prior assessments of practical skills by the students' own teachers were not very accurate. Since there is no significant tradition in the classroom of assessing practical laboratory skills for reporting or certification, it is likely that the majority of teachers have given little attention to deliberately and systematically assessing the laboratory skills of the own students. Indeed, the request from the trial organisers to supply such an assessment might well have been the first time that they ever considered it. In these circumstances, it would have been difficult for them to separate out a sound assessment of each student's practical skills from what they know about the student's achievements in general in the subject. One might expect that, in the event that a greater focus were to fall on the development and assessment of such practical skills in the future, this might well change.

Likewise, the correlation between the summer test scores submitted by teachers and the scores in the trial were not so high as to be worrying (0.45 for Chemistry, 0.42 for Biology, and 0.47 for Physics). If these were very high, it would suggest that the trial was measuring the same attribute as the teachers' written summer tests, which would not be a positive outcome. However, there are also many reasons why this correlation might be low, so we cannot conclude with any certainty from the data that we are accurately measuring something other than what the teachers' summer tests have measured. Nonetheless, in all three subjects, the supplied summer test marks correlated much more strongly with the booklet scores than they did with the directly observed performance scores. Given that we have established a comparatively high marking reliability in the latter, this is comforting, as it suggests that the directly observed performance scores are not only reliably measuring something, but that they are reliably measuring something that is different from what summer test scores measure.

In summary, while the analysis of the data does not provide strong convergent or discriminant evidence of the construct validity of the test, it does not undermine it either. The conclusion is that the validity argument must be predominantly developed on the basis of other forms of evidence and theory rather than this quantitative analysis of the trial data.

5.5 Fairness

While there are many aspects of fairness in the assessment that are dealt with elsewhere in this report, (and indeed many that are not addressed by the trial at all,) one aspect in particular is amenable to interrogation using the data. The nature of the assessment trialled involved different candidates being presented with different tasks. The question arises as to whether one can guarantee that these tasks are all of equal difficulty. If one cannot, and one were to roll out such an assessment as part of a high-stakes test, then this is a potential source of unfairness – two candidates of the same ability are not being treated equally if one is presented with a more difficult task than the other and no account taken of this in the scoring or grading.

This question of equivalence of task difficulty was explored using the trial data. Analysis of variance methods were used to test for significant differences in scores obtained across

different tasks. Similar methods were used to check whether the groups of students taking the different tasks were similar in prior achievement (as measured by the scores supplied by the teachers). While no statistically significant differences were found in the mean prior achievement of the students taking the different tasks, there were highly significant differences in the mean trial scores obtained across tasks (Table 5.7). That is, the students who did some tasks got significantly better marks than the students who did other tasks, even though they were no more competent at the subject than those other students. This is strong evidence of real differences in the difficulties of the different tasks.

Table 5.7 Evidence of significant differences in scores across task without significant differences in prior measure of candidate achievement

Subject	Test for differences in prior achievement across tasks	Test for differences in total trial score across tasks
Chemistry	$p = 0.58$ (not significant)	$p = 1.1 \times 10^{-5}$ (very significant)
Biology	$p = 0.09$ (not significant)	$p = 2.4 \times 10^{-16}$ (very significant)
Physics	$p = 0.36$ (not significant)	$p = 1.2 \times 10^{-9}$ (very significant)

To give a sense of the scale of variation in task difficulty, Table 5.8 shows the mean score for each task in each of the three subjects. The number of students that took each task and the standard deviations are also shown. The highest and lowest task-mean in each subject is highlighted for convenience. In all three subjects, the difference exceeds 20 marks.

Table 5.8 Candidate count, mean score (out of 120), and standard deviation for each task in each subject.

Task number	Biology			Chemistry			Physics		
	count	mean	s.d.	count	mean	s.d.	count	mean	s.d.
1	24	74.5	19.1	41	80.4	20.1	32	**104.5**	10.7
2	38	69.2	19.7	26	92.8	13.8	27	102.7	11.4
3	14	79.6	16.6	8	78.9	11.7	31	90.1	18.2
4	68	**89.9**	15.9	32	82.2	11.5	28	81.5	20.0
5	58	70.5	16.8	19	**74.0**	17.7	26	81.5	20.9
6	56	**63.8**	19.9	0	-	-	26	90.4	15.4
7	53	71.7	19.0	4	78.8	5.6	22	**75.7**	22.7
8	26	71.8	17.7	44	92.8	13.7	20	97.9	18.8
9	65	75.3	19.8	43	84.4	14.6	24	94.8	20.7
10	47	82.8	14.1	30	76.0	21.6	19	98.2	14.6
11	28	89.0	21.7	0	-	-	22	92.4	18.9
12	46	70.9	13.6	38	88.7	16.1	17	94.9	13.8
15				3	**94.7**	10.0			

This observed variation in task difficulty is not surprising. It is virtually impossible to guarantee equivalence of task difficulty in circumstances like these. Furthermore, even to strive for a very narrow range of task difficulty would impose a constraint on task design that would almost certainly narrow the intended test domain dramatically, as some of the

tasks that we wish to assess are simply intrinsically more challenging than others. Furthermore, it is well established that expert judgment (the only feasible means of establishing task difficulty in the absence of pre-testing) is not a sufficiently reliable method (Good and Cresswell 1988a; Cresswell, 1997; Cresswell, 2000; Baird and Scharaskin, 2002; Scharaskin and Baird, 2000). While variation in task difficulty might be considered tolerable in circumstances where candidates have a choice among tasks, it is much less acceptable in cases where the candidate has no such control. From a purely mathematical perspective, one can argue that a form of fairness is achieved by the fact that all candidates have the same probability of being assigned each task – whether hard or easy. This is not really satisfactory. It is 'fair' in the sense that everyone has the same 'luck of the draw', but this would not be regarded as fair in the ordinary sense of the word, and it certainly undermines validity, being a form of construct-irrelevant variance that can be avoided.

Accordingly, it should be recognised that fairness (and validity) would demand that, in the event of a full rollout of this model of assessment, differences in task difficulty would have to be accounted for by some method. The most reasonable approach would seem to be to apply suitable task-specific scaling transformations. This is not mirrored in any of the SEC's existing examination components, but it is not by any means an intractable difficulty. Subsequent performance in the written component could be used to confirm that the groups taking the different tasks shared similar achievement characteristics, following which the scores from each task could be re-scaled to a common distribution (such as the overall distribution across all tasks). Nevertheless, this would not absolve the task designers from striving to achieve as much comparability in difficulty as possible, both to minimise the amount of scaling required, and, more importantly, because of the potential negative psychological effect on candidates of having to engage with tasks that they find to be more difficult than those being done by others around them. Such effects would vary from person to person, and so cannot be fully addressed by scaling the results.

Chapter 6 Evaluating the technical quality of the assessment – qualitative evidence

While Chapter 5 interrogated the trial datasets to find objective quantitative evidence in support of or against the technical quality of the assessment, this chapter focuses more on the kinds of qualitative evidence that arise from the detailed application of expert judgment to data, observations and experience. It is based on the Chief Examiners' reports from the Examinations and Assessment managers (EAMs). Its sources of evidence are therefore the observations and judgments of the EAMs themselves, along with their interpretation and consolidation of the written reports completed by examiners following their work, and the verbal feedback from those examiners to the EAMs.

This chapter of necessity goes into subject-specific and task-specific detail, as it is only by these means that the thoughtful judgment of subject experts can be brought to bear. Importance evidence for or against validity arises from the degree to which subject experts are satisfied that the assessment is measuring the right skills and that its content is soundly based on the important concepts, principles and practices of the discipline. The assessment of *practical skill*s by direct observation in each subject is treated first, followed by the assessment of *results and analysis* as captured by the work in the booklets.

In each case, the analysis begins with data and commentary on how the students performed on the specific tasks involved, following which some overall observations arising from this analysis are made.

While this chapter does contain some statistical information, this is primarily provided to contextualise, justify or illuminate the observations being made.

6.1 Assessment of *practical skills* – Biology

6.1.1 Analysis by task and by assessment objective of the practical skills marks awarded

Data

Table 6.1: Number of times each task was trialled, and average *practical skills* mark for each task – Biology

Task no.	No. students	% of students	Average mark (%)	Rank order by average marks	Task Topic
1	24	4.6	88.2	1	Photosynthesis rate vs. CO_2 concentration
2	38	7.3	80.7	7	Photosynthesis rate vs. light intensity
3	14	2.7	81.0	= 5	Effect of IAA on plant tissue

4	68	13.0	79.3	8	o Examining plant cells with a microscope o Examining animal cells with a microscope
5	58	11.1	72.8	12	o Qualitative food tests o Enzyme denaturation
6	56	10.7	75.7	10	Enzyme activity vs. pH
7	53	10.1	81.8	= 3	Enzyme activity vs. temperature
8	26	5	75.3	11	Plant tissue DNA isolation
9	65	12.4	78.5	9	Effect of solute concentration on water movement across a selectively-permeable membrane
10	47	9.0	81.0	= 5	Mammalian heart dissection
11	28	5.4	81.8	= 3	o Animal population quantitative survey o Seed germination conditions
12	46	8.8	82.5	2	Growing leaf yeast

Table 6.2: Average mark awarded under each *assessment objective* heading – Biology

Assessment Objective	Average mark (%)
Selection of apparatus and other materials	89.2
Assembly of apparatus	81.7
Use of apparatus	71.7
Observations / measurements	72.5
Working safely and efficiently and cleaning up	84.2

Commentary

The tasks were written for a common level standard so, given that about three-quarters of Biology candidates in the Leaving Certificate examination take Higher Level, most students exhibited a high level of achievement.

Selection of apparatus, chemicals and other materials

Students did very well in selecting the correct apparatus, chemicals and materials for their tasks. Some students in some schools were not familiar with the names of items e.g. boiling tubes, mass balance, cork borer, catalase source. A small minority could not identify glass slides and coverslips.

Assembly of apparatus and preparation of biological samples

Examiners generally reported that the assembly of apparatus and preparation of biological samples were well performed. Difficulties experienced by students mentioned in many examiners' reports were:

- In centres where thermostatically controlled waterbaths were not available, many students demonstrated unfamiliarity with the assembly and preparation of alternatives.
- Lighting a Bunsen burner and regulating the flame proved problematical for some students, to the extent of unsafe practice in some cases.
- Many students did not label solutions once they had collected them, which had the potential to cause confusion where more than one solution was being used in a given task.
- Some students did not add water to the pieces of celery before blending when preparing to extract catalase in tasks B5.2, B6 and B7. If a blender was not used, many students did not chop the celery into small enough pieces, some using inappropriately large chunks. Similarly, some students did not seem to be familiar with how to find and remove the epidermis from onion leaves.
- Elodea stalks were often cut in air before immersion in $NaHCO_3$ solution.
- Most students seemed unfamiliar with cork borers for cutting discs from the leaves in the leaf yeast task.

Use of apparatus

Most students handled equipment safely; there were few spillages or breakages. Almost all students worked more quickly than they needed to, which often led to examiners missing crucial steps in tasks. Tasks with more comprehensive instructions were generally better performed although sometimes the detailed instructions were not followed, either through being ignored, not being read properly, students not understanding certain terminology, or through rushing.

Examiners reported that some students were unfamiliar with:

- the correct procedure for focusing a microscope;
- the use of dividers when measuring the thickness of the heart chamber walls;
- the importance of and adherence to aseptic technique;
- the necessity of heating Benedict's solution when testing for reducing sugars;
- the necessity to use a new or a clean dropper or syringe with each new solution;
- methods of controlling the temperature of non-thermostatically controlled waterbaths.

Examiners also reported that students had difficulty manipulating the *Visking* tubing in task B9 and measuring volumes accurately.

Observations / measurements

Most students understood what was to be observed or measured and knew the required technique but the execution was not always competent.

Observations and measurements in the enzyme tasks B6 and B7 were reported as having been well done overall, although the use of pH meters was rare, and not of a very high standard when employed. Generally, use of the electronic mass balances was good.

While most students measured volume accurately, a significant minority did not make sure the graduated vessel was on an even surface before attempting to measure volume at the meniscus. Some did not know to read volume at the meniscus, and incorrect reading of instructions sometimes led to incorrect volumes being measured. Volume measuring with Pasteur pipettes was reported as poor, with students failing to ensure the absence of bubbles from the liquid in the pipette.

Students were occasionally observed removing the thermometer from the solution in order to measure the temperature of the solution. Examiners reported that students were frequently careless about making sure that the time intervals between successive measurements were equal, as specified in the procedure.

There was some evidence of poor calculations in completing the tables in B6 and B7. For instance, some candidates proved unable to arrive at a correct figure for 'volume of foam produced' by subtracting the initial volume from the final volume. Similarly, in task B1, there was some evidence of inability to calculate the average number of bubbles produced from the three attempts at each CO_2 concentration.

Examiners encountered very few instances of students repeating a task in order to generate confirmatory data. In certain tasks (B1, B2, B6, B7, B8, and especially B5) this was at least partly explained by the time-consuming nature of the task.

Working safely & efficiently & cleaning up

Most students worked safely in a systematic and efficient manner. Few spillages or breakages were reported and those that did occur were quickly cleaned up by the student.

Chemicals were generally used economically although in some cases if a student used an inappropriately large container to collect a particular chemical e.g. a beaker instead of a test tube, the volume of chemical collected was too large. This practice led to shortages of some chemicals in some schools.

Most students worked systematically through the instructions in the procedure section of their task, with those tasks having the most comprehensive instructions being most efficiently carried out. A minority of students, however, seemed to have largely ignored the instructions and carried out the task as they remembered having done it in laboratory class.

Students generally moved carefully around the lab and for the most part worked quietly and efficiently.

The most commonly reported unsafe practice was students not wearing safety goggles or gloves when handling the hydrogen peroxide in tasks B5.2, B6 and B7; and when handling the IAA solutions in task B3. Other unsafe practices were long hair not being tied back, some students not washing the heart before dissecting; having electric leads too close to hotplates; overfilling Bunsen burner-heated waterbaths; using the scalpel unsafely.

The majority of students kept their workstations clean and tidied up as they went along. The messiest practice observed was where students cut plant material on the lab bench surface instead of using a chopping board. The final tidy-up was done to a high standard with almost

all workstations being left in good order. Just one school was reported in which the students were not good at cleaning and tidying up after themselves.

The Biology tasks were not all of equal demand in terms of time needed, with some students finishing up to 30 minutes earlier than those who were assigned the more time-consuming tasks. Task B5 was reported by all advising examiners as being the most difficult for students to finish on time.

6.2 Assessment of *practical skills* – Chemistry

6.2.1 Analysis by task and by assessment objective of the practical skills marks awarded

Data

Table 6.3: Number of times each Chemistry task was trialled, and average *practical skills* mark for each task

Task no.	No. students	% of students	Average mark (%)	Rank order of average marks	Task Topic
1	41	14.2	80.5	7, 8	Flame test/copper Na_2CO_3 primary std & titration
2	26	9.0	89.7	2	Salt identification Heat of neutralisation
3	8	2.8	79.0	10	Rate of reaction EDTA titration
4	32	11.1	84.3	6	Fe^{2+} and $KMnO_4$ titration Salt anion identification
5	19	6.6	80.5	7, 8	Prep and tests ethyne Test for chloride ion
6	Task not used in trial				Melting Point Comparator / DPD
7	4	1.4	78.3	11	Benzoic acid recrystallisation Diluted vinegar titration
8	44	15.3	88.3	4	Rate, Temp Test for carbonate ion
9	43	14.9	87.3	5	Unknown salt (nitrate) Rate, concentration
10	30	10.4	80.3	9	Rate, gas evolution Standard solution Na_2CO_3
11	Task not used in trial				Prepare Benzoic Acid HCl/NaOH titration
12	38	13.2	90.0	1	$PV = nRT$ Vinegar dilution
15	3	1.0	89.2	3	Test for chloride ion Diluted vinegar titration

Task 15 in Table 6.3 above was an additional task produced at a later stage of the trial so that too many students would not be doing the same tasks in some sessions.

Table 6.4: Average mark awarded under each *assessment objective* heading – Chemistry

Assessment Objective	Average mark (%)
Selection of apparatus and other materials	90.0
Assembly of apparatus	90.0
Use of apparatus	81.7
Observations / measurements	80.8
Working safely and efficiently and cleaning up	85.8

Commentary

The Chemistry tasks were set at a common level standard. . It is noted that of the order of 85% of students present at Higher level in this subject. Good performance in the tasks required that the students could follow the instructions given in the task booklet, and had a familiarity with locating and using apparatus in the laboratory and competence in performing certain techniques.

Selection of apparatus, chemicals and other materials

Students scored well in selecting the correct equipment, chemicals and materials for their tasks. Some students however used beakers where conical flasks would have been more appropriate, and some were unaware that the graduations on conical flasks and beakers are not suitable for accurate measurements.

Assembly of apparatus

Most students showed good competence in arranging and assembling titration apparatus. Students experienced problems with the assembly of the gas preparation apparatus in C5 and C10. Some students were unable to use clamps correctly to help support the apparatus and a few had to be instructed to fill the collection vessel with water.

Use of apparatus

Most students were able to use the apparatus available to complete their task. However, classic errors in the technique when using titration apparatus were observed, e.g. a funnel left in a burette after filling and when adjusting to zero. Wash bottles were not used in titrations by some students. Students sometimes used graduated cylinders for measuring volume when volumetric flasks would have been appropriate.

Thermometers were often observed in reaction flasks instead of in water baths. Some students had difficulties carrying out the vacuum filtration in task C7. Many students had to be guided, for safety, to use certain chemicals in the fume-hood and very few were able to switch on the fume-hood themselves.

The collection of oxygen in C10 was problematic for a number of students. Students who had previously performed this experiment with a partner found it difficult to perform alone even though only a small number of readings of time and volume were required. The poor quality of the hydrogen peroxide supplied in some schools was another frustration for some students.

Competent use of the apparatus in measuring the relative molecular mass of a volatile liquid in C12 and in carrying out the serial dilutions in C9 was observed. Some students had difficulty lighting and adjusting a Bunsen burner flame. Some students did not turn off the Bunsen flame immediately after use.

Observations / measurements

Good use of balances by most students was recorded. Some students read the mass correctly at the balance and then recorded an incorrect mass in their task booklet at their bench.

Meniscuses were not always read at eye-level although most students were observed to have used great care with droppers to bring the meniscus correctly to the calibration mark in volumetric flasks and in some cases, unusually, burettes. Some students had difficulty in getting the pipette reading exactly on the calibration. In general, students read end points correctly and recorded flame colours correctly. Several examiners commented on students reading the volume in graduated cylinders, volumetric flasks, etc held at eye-level in the hand.

Working safely & efficiently & cleaning up

The majority of students acted safely and efficiently and completed their tasks in the time allowed although some did not leave sufficient time to complete their task booklet. Some were still tidying their workstation right up until the end of the session and a few had insufficient time for the clean-up. As the tasks were of equal demand in terms of time, this meant that students were all busy and on their feet throughout the 90-minute session. The students worked quietly with minimal interaction between them and moved around the laboratory at an appropriate speed. Most students completed the task booklets at the end while it may have been more efficient to partially complete them as they proceeded through their tasks. Several students requested graph paper in the last ten minutes of the session.

Long hair was tied back. Students wore safety glasses. Students also had a tendency to remove the safety glasses when completing the task booklets towards the end. Students had to be prompted to work in the fume-hood and about wearing gloves when using concentrated acids. Students needed to guard against wet glassware slipping out of their hands.

Relatively few students labelled the beakers, flasks and the other containers they used to complete their tasks. Many students overreached or stood on their toes or on furniture to fill burettes instead of using the retort stand and clamp correctly. Forcing pipette fillers onto pipettes and holding the pipette filler and pipette end too far apart during the fitting on of the pipette filler were other unsafe practices observed.

6.3 Assessment of *practical skills* – Physics

6.3.1 Analysis by task and by assessment objective of the practical skills marks awarded

Data

Table 6.5: Number of times each Physics task was trialled, and average *practical skills* mark for each task

Task no.	No. students	% of students	Average mark (%)	Rank order of average marks	Task Topic
1	38	13.0	86.3	8	Converging lens
2	35	12.0	87.2	5	Curved mirror
3	27	9.2	86.8	6	Refractive index
4	17	5.8	78.7	11	Specific heat capacity of a liquid
5	19	6.5	76.5	12	Specific latent heat of a substance
6	22	7.5	87.5	4	To investigate variation in the thermometric property of a material with temperature
7	27	9.2	85.2	10	To measure *g*, the acceleration due to gravity by free fall
8	21	7.2	88.0	3	To measure the resistivity of the material of a wire
9	26	8.9	90.8	1	To investigate the laws of equilibrium for a set of co-planar forces
10	25	8.6	89.7	2	Boyle's law
11	21	7.2	86.7	7	To verify that acceleration is proportional to force
12	14	4.8	85.3	9	To verify the principle of conservation of momentum

Table 6.6: Average mark awarded under each *assessment objective* heading – Physics

Assessment Objective	Average mark (%)
Selection of apparatus and other materials	90.8
Assembly of apparatus	90.0
Use of apparatus	78.3
Observations / measurements	75.8
Working safely and efficiently and cleaning up	94.2

Commentary

The Physics tasks were set at a common level standard. Good performance in the tasks required that the students could follow the instructions given in the task booklet, and had a familiarity with locating and using apparatus in the laboratory and competence in performing certain techniques.

Selection of apparatus, chemicals & other materials
Students succeeded in selecting the correct equipment, chemicals and materials for their tasks and most scored very well.

Assembly of apparatus
Examiners reported very good competence by students in arranging and assembling apparatus. Students had problems with assembling the apparatus in P4 and P5, which concerned heat. Some students had problems with tasks P11 and P12, depending on whether or not light-gates were being used. Some students found the setup of the one electricity task (P8) to be quite difficult.

Use of apparatus
Technical errors were found in a number of tasks; errors by students in taring mass balances and errors of parallax in reading metre sticks and thermometers were common. Tasks P11 and P12 were problematic for students, depending on the equipment used. Tasks where datalogging equipment was used such that the students were not taking direct measurements themselves, or then manipulating this data also caused problems for students. Some students found the use of the micrometer and the multimeter in task P8 to be a challenge.

Observations / measurements
Examiners reported some errors in reading scales on thermometers and metre-sticks. In the case of tasks P1 and P2, errors related to object and image distances were observed. In the case of task P3, a large number of students displayed confusion regarding what angles to measure. Most mechanics experiments did not cause difficulties, where students were familiar with the equipment they were using.

Working safely and efficiently & cleaning up
The majority of students acted safely and efficiently and completed their tasks in the time allowed and had plenty time to complete their booklets. Not all students were always busy and on their feet throughout the 90-minute session, which indicated the possible need for the introduction of sub-tasks in the Physics tasks. The students worked quietly with minimal interaction. It was noticeable that students waited until near the end to complete the task booklets even though it might have been more efficient to partially complete them as they proceeded through their tasks.

Some Physics students requested white coats and laboratory glasses but many did not. While Physics tasks may not lead to the same safety concerns as those found in Chemistry and Biology, students carrying out tasks P4, P5 and P6 had to be aware of safety issues arising from the use of hot plates or similar.

6.4 Some overall observations on tasks and performance

- Students working quickly at Biology tasks led to difficulties as examiners missed important steps in tasks.
- Some unsafe practices were observed in Biology and in Chemistry.
- Some tasks in Biology were finished up to 30 minutes earlier than others.
- Chemistry tasks had an equal time demand.
- The time demand of Physics tasks varied significantly.
- Many of the errors or deficiencies observed in the practical skills displayed by students in the trial are consistent with poor or limited exposure and training in practical laboratory skills as part of the current delivery of the subject.

6.5 Assessment of *results and analysis* – Biology

6.5.1 Analysis by task of the *results and analysis* marks awarded

Data

Table 6.7: Number of times each Biology task was trialled, and average *results and analysis* mark for each task

Task no.	No. of students	% of students	Average mark (%)	Rank order by average marks	Task Topic
1	24	4.6	36.0	9	Photosynthesis rate vs. CO_2 concentration
2	38	7.3	34.7	11	Photosynthesis rate vs. light intensity
3	14	2.7	51.7	4	Effect of IAA on plant tissue
4	68	13.0	70.7	1	Examining plant cells with a microscope Examining animal cells with a microscope
5	58	11.1	44.7	5	Qualitative food tests Enzyme denaturation
6	56	10.7	30.8	12	Enzyme activity vs. pH
7	53	10.1	37.7	8	Enzyme activity vs. temperature
8	26	5.0	44.5	6	Plant tissue DNA isolation
9	65	12.4	39.5	7	Effect of solute concentration on water movement across a selectively-permeable membrane
10	47	9.0	57.2	3	Mammalian heart dissection

11	28	5.4	66.3	2	Animal population quantitative survey Seed germination conditions
12	46	8.8	35.7	10	Growing leaf yeast

Commentary

There was a wide variation in the quality of student answering in the Biology tasks, with marks for the *results and analysis* varying from 0% to 95%. As examiners marked between 32 and 34 task booklets, the number of any particular task booklet marked by each examiner was low. Because of the small number of booklets involved for each, examiners' general impressions were sometimes contradictory. That is, almost every task was described by some examiners as having been well done, while by others as average or even poor.

There was a big variation in the quality of students' answers to questions seeking explanations for particular steps in procedures. This reflected the range of ability of the students taking part in the trial, and the fact that the trial was at a common level. Examiners reported that many students appeared to have an incomplete grasp of terms such as *reliability*, *significance*, *uncertainty*, *limiting factor*, and *variables.* A significant minority of students seemed not to be used to graphing their results, and as a result they made many elementary mistakes in labelling and scaling of the axes.

6.6 Assessment of *results and analysis* – Chemistry

6.6.1 Analysis by task of the *results and analysis* marks awarded

Data

Table 6.8: Number of times each Chemistry task was trialled, and average *results and analysis* mark for each task

Task no.	No. of students	% of students	Average mark (%)	Rank order by average marks	Task Topic
1	41	14.2	53.3	5, 6	Flame test/copper Na_2CO_3 primary std & titration
2	26	9.0	65.0	3	Salt identification Heat of neutralisation
3	8	2.8	52.5	9	Rate of reaction EDTA titration
4	32	11.1	52.7	8	Fe^{2+} and $KMnO_4$ titration Salt anion identification
5	19	6.6	42.8	11	Prep and tests ethyne Test for chloride ion
6	Task not used in trial				Melting Point Comparator / DPD
7	4	1.4	53.0	7	Benzoic acid recrystallisation Diluted vinegar titration

8	44	15.3	66.5	2	Rate, Temp Test for carbonate ion
9	43	14.9	53.3	5, 6	Unknown salt (nitrate) Rate, concentration
10	30	10.4	46.3	10	Rate, gas evolution Standard solution Na_2CO_3
11	Task not used in trial				Prepare Benzoic Acid HCl/NaOH titration
12	38	13.2	58.0	4	$PV = nRT$ Vinegar dilution
15	3	1.0	68.8	1	Test for chloride ion Diluted vinegar titration

Commentary

There was a wide variation in the quality of student answering with marks for the indirect assessment varying from 10% to 99%. As each of the nine examiners marked just 32 task booklets, the number of any one task marked by each examiner was in the small single figures. As a result, it was difficult to obtain an accurate evaluation from examiners' reports alone on the quality of students' answering in the case of most tasks. However the student task booklets revealed that some students who were able to carry out a practical task very well, and scored well in direct assessment, were unable to correctly explain in their booklet the chemical rationale behind some of the procedures they had followed. Similarly, some students were unable to perform the calculations associated with the practical task without errors and consequently scored poorly in indirect assessment. Given the wide range of student ability represented in the trial and the stage of the Leaving Certificate programme at which the trial took place this is not surprising.

Several examiners commented on the fact that students had not completed the task booklets. This could have been a time-management issue on the part of the students, or it may indicate that some of the Chemistry tasks were too long.

6.7 Assessment of *results and analysis* – Physics

6.7.1 Analysis by task of the *results and analysis* marks awarded

Data

Table 6.9: Number of times each Physics task was trialled, and average *results and analysis* mark for each task

Task no.	No. students	% of students	Average mark (%)	Rank order of average marks	Task Topic
1	38	13.0	49.7	9	Converging lens
2	35	12.0	50.3	8	Curved mirror
3	27	9.2	42.7	11	Refractive index
4	17	5.8	42.2	12	Specific heat capacity of a liquid
5	19	6.5	43.5	10	Specific latent heat of a substance
6	22	7.5	57.0	3	To investigate variation in the thermometric property of a material with temperature
7	27	9.2	53.5	5	To measure *g*, the acceleration due to gravity by free fall
8	21	7.2	55.2	4	To measure the resistivity of the material of a wire
9	26	8.9	61.2	1	To investigate the laws of equilibrium for a set of co-planar forces
10	25	8.6	60.5	2	Boyle's law
11	21	7.2	52.7	7	To verify that acceleration is proportional to force
12	14	4.8	53.3	6	To verify the principle of conservation of momentum

Commentary

There was a wide variation in the quality of student answering with marks for the indirect assessment varying from 5% to 100%. Each examiner marked approximately 36 task booklets, and therefore the number of booklets marked by each examiner was small. For this reason, examiner impressions were often contradictory, e.g. some examiners reported students having excellent understanding of percentage error and others reported that this was a source of difficulty for students.

Some students who were able to carry out a practical task very well, and scored well in direct assessment, struggled with the associated calculations and with explaining the rationale governing the procedures followed and scored poorly in indirect assessment. The design of the tasks for common level meant that certain items known to be very challenging to Ordinary Level students, e.g. drawing graphs, using graphs, manipulating and using mathematical formulae were poorly answered by a significant minority of the trial cohort.

Examiners commented that the work of most students did not indicate that time was a factor in any failure by students to properly complete the task booklets.

6.8 Issues that affected implementation of the practical assessment in the trial

This section summarises a range of factors that led to difficulties on the part of students in completing the practical examination, as reported by examiners and EAMs. The issues that arose have been categorised into student issues and school issues, even though they are usually interrelated.

6.8.1 Student issues

Biology

- In some cases students collected beakers to carry out the task, where test tubes were the correct choice, or graduated cylinders where beakers were intended, and this sometimes led to a shortage of beakers or graduated cylinders for other students.
- In general, chemicals were supplied by schools in clearly labelled containers and this allowed for easy recognition and collection. However, in cases where beakers had been inappropriately chosen by students rather than test tubes, large aliquots of particular chemicals were collected in these beakers, leading to shortages elsewhere.
- *Task B4 (Examining plant cells and animal cells with a microscope):* Examiners reported that the × 40 objective lenses in laboratories were often of inferior quality and did not produce good quality images. While poor quality lenses may have been the cause of this, examiners also reported that some students' poor skills at microscope manipulation led to × 40 objectives being so carelessly used that they were driven through the coverslip/ slide on the stage.
- *Task B10 (Mammalian heart dissection):* Many students, on learning they had been assigned the heart dissection, elected not to take it, either through squeamishness or conscientious objection.

Chemistry

- The collection of oxygen in Task C10 was problematic as students who had previously performed this experiment with a partner found it difficult to perform alone despite the fact that only a small number of readings of time and volume were required.

Physics

- Many students were unaware of how to find the approximate focal length of a mirror/lens at the beginning of an experiment.
- In the case of the experiments on motion, the fact that students had to work alone led to them having difficulty in setting up the experiments and taking readings.

6.8.2 School issues

Biology

- In most schools, the equipment to be used in the examination was made available by the designated support teacher (DST) in a single area in the laboratory. The equipment was most often made available on a trolley or set of trolleys, or sometimes on a bench at the back or front or side of the lab.
- There were no reports of tasks being impossible to do because of lack of equipment or chemicals although some schools were short of some items of equipment such as thermostatically controlled water baths and accurate mass balances. In some laboratories, students had to queue to use an accurate balance. This could easily be addressed if schools acquired a number of modern, cheap, jeweller's balances.
- A problem outside the students' control was that in many centres the hydrogen peroxide supplied by schools for tasks B5.2 , B6 and B7 either was supplied at an incorrect concentration or was of such poor quality as to be useless.
- The most widespread problem outside the students' control was the poor condition of the pondweed *Elodea canadiensis* that was needed for the two photosynthesis tasks B1 and B2. In some centres examiners reported that the DST had not prepared the *Elodea* by illuminating it for a period before the start of the task sessions. Some DSTs supplied *Cabomba* spp. instead, which seemed to work better but *Cabomba* is an invasive plant the use of which is discouraged for that reason.
- *Task B1 (Photosynthesis rate vs. CO_2 concentration):* The main issue here was the poor quality of the supplied *Elodea* pondweed. In the great majority of cases, students who attempted to perform this task had to ask for pseudo data.
- *Task B2 (Photosynthesis rate vs. light intensity):* Here again the main issue was the poor quality of the supplied *Elodea* pondweed. Similarly, to B1, in the great majority of cases students who attempted to perform this task had to ask for pseudo data.
- *Task B4 (Examining plant cells and animal cells with a microscope):* Examiners reported that the × 40 objective lenses were often of inferior quality and did not give good quality images.
- *Task B5 (Qualitative food tests; Enzyme denaturation):* Biuret reagent was not given in the equipment list for sub-task 5.1. Instead, the constituents of biuret solution ($CuSO_4$ and NaOH) were listed and this led to confusion, as most schools now use already formulated biuret compound.
- In sub-task 5.2 the procedure was included for doing the investigation using catalase. In a small number of schools, students had done this investigation using amylase and were not familiar with the details of the catalase procedure. In addition, while celery and radish were the catalase sources given in the DSTs' equipment list, some students were used to using liver for this purpose, which, for various reasons, was unsuitable in the context of the trial.
- In sub-task 5.2, the hydrogen peroxide was either supplied at the wrong concentration (20% was specified in the equipment list but the DST could have mistaken this for 20

vol., which is equivalent to 6%) or was old or had been inappropriately stored and had already at least partially decomposed and lost its potency.

- *Task B6 (Enzyme activity vs. pH):* pH meters supplied for use in this task were usually uncalibrated, and students were not familiar with their operation in most cases.

Chemistry

- Students were able to find the equipment in its usual location in the laboratory or where the trial took place in a laboratory unfamiliar to them, it was easiest for them when the equipment was laid out on benches for the practical assessment.
- Examiners reported no issues about tasks being impossible to do because of lack of equipment or chemicals although some improvisation did occur. In some laboratories, students had to queue to use the accurate balance(s). This could be addressed through the acquisition by schools of modern, cheap, jeweller's balances.
- Count-up timers were in short supply in some schools.
- Some schools had shortages of suitable gas-preparation glassware, Buchner flasks and Hirsch funnels.
- Not all laboratories had deionised water and, in some cases, the deionised water was bought into the laboratory in containers instead of being deionised *in situ*.
- White coats were not available in all schools and plastic aprons, as used in some sessions, were not ideal given the use of Bunsen burners in some tasks.

Physics

- Some schools found it difficult to carry out both task P11 and task P12 because of a lack of availability of sufficient mechanics equipment.

6.9 Some overall observations and suggestions arising from the issues that arose

Observations

- Many examiners reported students feeling under pressure because they were performing their task alone rather than in the pair or group scenarios more familiar to them from classroom experience.
- White coats were available in most schools, and where available were always worn. A minority of schools supplied aprons instead of white coats. The aprons were thought by examiners to be unsatisfactory as they were generally plastic coated and too short.
- There was inconsistency in the labelling in the trial and instances were observed of students selecting the correct chemical but of the incorrect concentration for their task.
- There was interaction between the students and the examiners in some cases to locate the chemicals required for tasks.

- If more of the examination tasks included the preparation of stock solutions from solids, this would reduce the number of solutions that had to be prepared for the trial.
- The inability of students to switch on the fume-hood for themselves and their obvious unfamiliarity with working in one suggests that fume-hoods are seldom used by the students themselves as a routine part of their practical work.

Suggestions

- Examiners reported on the difficulties they experienced when students had proceeded past a critical measurement in a task. To address this, they recommended that each task should contain a small number of steps where the student is required to call the examiner to check a measurement. Examiners also advised the generation and use of task-specific lists of observable skills and measurements for each examiner's use in direct assessment.
- It was recommended by examiners that apparatus should be left in its usual location in the laboratory but the examiner should check before the sessions that sufficient equipment was available. Where equipment is stored out of sight in cupboards or drawers, labels should indicate the contents of the cupboard or drawer. Students should be assessed where they usually perform practical work, where possible.
- Separate locations need to be designated in laboratories for (i) standard reagents such as 3 M hydrochloric acid, (ii) indicators, (ii) solid samples required for tasks, and (iii) acids and bases required for tasks.
- All advising examiners were of the opinion that it would have been very useful if examiners had been allowed to ask questions of the students as they observed their work.
- Some examiners and advising examiners recommended that tasks should directly examine students' proficiency at performing certain generic skills such as measuring liquid volumes, filling pipettes, using a balance, preparing a microscope slide for examination, separately from the context of a particular investigation.
- Some Biology examiners and advising examiners decried what they saw as the overly prescriptive nature of the instructions in the tasks and advocated the assessment of students' ability to come up with practical solutions to biological problems based on the skills they had learned during their laboratory work for the Leaving Certificate course.
- If practical assessment in Chemistry were to be implemented, it was suggested that schools would need guidance from the SEC regarding containers sizes and types. A standard set of labels could be issued by the SEC to schools by e-mail (to be printed on sticky labels in the school). SEC guidance about the layout of chemicals in the laboratory would also be helpful for candidates, schools and examiners, and would help to achieve a more consistent experience across different practical examination sessions.

6.10 Some key messages from the chief examiners' reports

General

The experience of the examiners and of the EAMs and members of the implementation group who visited schools during the trial, was that students', teachers' and school managements' reaction to practical assessment and to the model trialled were very positive.

There is general agreement that teaching and learning would be enhanced by the introduction of practical assessment and that students would have greater sense of involvement in and ownership of practical work. It would emphasise scientific enquiry, skills acquisition and purposeful science practice in science education. Practical assessment would reward students who engage well with practical work and acquire good practical skills. It would help students to acquire transferable skills of observation and measurement, manual dexterity, and a good attitude to safety, based on a balanced approach to hazards and their associated risks, which can be applied outside the laboratory.

In the case of Chemistry and Physics, some practical work specified as Higher Level in the current syllabus is designated common to both levels in the draft subject specifications on which practical assessment in the future will be based. Consideration will need to be given to the means by which the assessment can give fair and reasonable results to both cohorts, whether that be through the application of different marking schemes to the marking of the booklets at the two levels, or by other means.

The draft Physics specification allows a wide variety of methods for carrying out many of the practical activities, and these methods can be of widely varying levels of difficulty to implement. Unless this is addressed, different students given the same overall task will be carrying out significantly different experiments, using different techniques and different equipment, and which include calculations of widely varying degrees of difficulty. This raises issues of fairness that are difficult to resolve.

Irrespective of whether data-logging equipment is used for measuring, it would be problematic from an equity perspective if such equipment and/or computer software were to be used to perform calculations or analysis as well. Until such time as all students have similar levels of access to such equipment as a matter of routine, all students should be required to analyse their data and perform relevant calculations without the assistance of such technologies.

Designated support teacher (DST) and laboratory preparation

DSTs in the trial engaged very positively with examiners and were supportive of the role of the examiner. Their support of practical assessment before and during the assessment is critical.

The introduction of practical assessment would require a planned approach by schools' science departments to laboratory and storeroom management and ordering of equipment. Appropriate continuous professional development for teachers to support this will need to be made available.

The volumes of solutions specified in Biology and Chemistry tasks should be minimised for environmental reasons, to economise on use of chemicals, to save storage space, and to reduce deionised water consumption.

Preparation of solutions for the Chemistry trial took DSTs on average about 20 hours. For the assessment to be feasible in practice, this would need to be reduced. Possible means of doing so include: schools having containers of suitable sizes and sets of standard labels; ensuring that more tasks incorporated student preparation of stock solutions; and making available a suitable list of reagents in schools from year to year for use in lessons and in the assessment. Continuous professional development in bench Chemistry safety and chemical storage would complement this.

Examining

As part of the preparation of the practical assessment, the tasks should be tested by setters, and examiner training should involve a practice session to familiarise examiners with the tasks. To minimise the possibility of a systems failure in a school during the practical examination, examiners should check that the necessary equipment is available and in working order, and that solutions were correctly prepared and available in sufficient quantity. This would require that the examiners have a time allocation and receive remuneration for this.

Consideration should be given to providing examiners with a list, specific to each task, of the observable skills in that task and of the opportunities in the task to assess students' skills of measurement.

A student from a non-exam class acting as a helper (to call the DST) during examining sessions, as was recommended, was found to be necessary during the trial.

During the trial it was generally found possible to assess 12 students in a session, with some difficulty. As this difficulty was increased where one or more students required help, examiner guidelines should provide guidance on the maximum time that could be spent helping a student. It might be wiser to reduce the maximum number of students somewhat.

Students in the Biology or Chemistry practical examination should be required to label any vessel into which they place a chemical to assist the examiner in assessing their practical work.

6.11 Reflecting on the tasks in light of implementation the trial

It is appropriate to reflect all aspects of the practical examination tasks in order to provide guidance in setting tasks in any future practical examination. This section examines the differences in structure, implementation, and student responses to the tasks in the three subjects. The purpose of this is to highlight commonalities and differences in the experience of implementing the trial across the three subjects, so that future implementation in each might benefit from the others.

6.11.1 Nature of the tasks

The tasks were largely based on prescribed practical activities in the current syllabuses that are common to the new specifications. Each task was based on one or more of the practical activities. The task topics are indicated in the tables earlier in this chapter, and the full set of task booklets used in the trial are presented in Appendix E (Biology), Appendix F (Chemistry) and Appendix G (Physics).

Students were assessed on their practical skills in accordance with the new specification for each subject. The practical skills assessed were grouped under the following headings: selecting the appropriate apparatus, chemicals and other materials; working safely and efficiently; assembling apparatus; using apparatus; making observations and measurements; and working safely and efficiently. Each task included questions on the data collected by students. The questions also tested students on their observations and conclusions and on their ability to analyse, report on and communicate their results in an appropriate scientific manner.

The setters' guidelines for generating the tasks are in Appendix C. The assessment of practical skills by direct observation is an authentic form of assessment in that it does not rely on secondary or indirect evidence of practical skills, but is a direct assessment of the relevant skills being displayed by students in their own laboratory environment, carrying out activities that closely mirror the intended learning activities in intended by the specification, and, by and large following a complete experimental cycle (as distinct from testing constituent skills in isolation). To some degree, the skills assessed through the work in the booklet are skills that can also be assessed in a written examination. However, the capacity to apply those skills in a 'live' context to primary data that the students have generated themselves through their own practical experimental work is a distinct competence that does not necessarily follow from their capacity to apply them in a written examination to data provided by someone else. The assessment of the booklet therefore contributes significantly to the authenticity and validity of the assessment component as a whole.

Notwithstanding these positive features, the practical assessment proposed and trialled is, of necessity, an individual assessment of a student working independently, and this might be considered a source of *inauthenticity*, since the normal practice in science practical work is to work in groups of two or three students. Also, as specific tasks are developed in the cold light of the knowledge of the different contexts, equipment and methodologies that arise in different classrooms across the country, and the need not to disadvantage candidates on the basis of differences in any of these factors, tasks can end up being more constrained in what they can achieve than might initially be hoped for.

6.11.2 Comparison of tasks between subjects

In the case of the Physics and the Biology tasks, almost all tasks were based on just one topic. For Chemistry, each task contained two sub-tasks, based on different unrelated topics.

Having two sub-topics in a task made it easier to have the tasks of roughly the same length. This was borne out by the observations of examiners and visitors to the trial who commented on the fact that students in the Chemistry trial all seemed to require the full 90 minutes of examination time to complete their tasks whereas for Physics and Biology there were big variations.

The level of challenge for students of the Chemistry tasks was reported to be more consistent between tasks compared to the Biology and Physics tasks, where large differences in the challenge level of tasks were reported. If this held true, then it would mean that in the case of Physics and Biology, students could be lucky or unlucky depending on the particular task they were assigned. However, notwithstanding the more favourable reports in this regard in

relation to Chemistry, data analysis indicated clearly that there were significant differences in task difficulty in all subjects, and this issue and how it might be addressed is dealt with more fully in section 5.5 of Chapter 5.

6.11.3 Structure of tasks

Some characteristics of the tasks in each subject are outlined below.

Biology tasks

- Each task was based on a single experiment and in three cases the tasks were divided into two sub-tasks. The title of the experiment was stated in a box at the beginning of the task.
- Each task or sub-task commenced with a brief introduction of the background theory to the task.
- The introduction to each task was followed by a list of the apparatus and the reagents that were required to carry out the task.
- Detailed step by step instructions were outlined for carrying out the practical work in each task.
- Some of the tasks appeared to be short.

Chemistry tasks

- Each of the Chemistry tasks was composed of two sub-tasks. The title of each sub-task was stated in a box at the beginning of the sub-task.
- In almost all tasks, one of the sub-tasks was much shorter than the other.
- The directly assessed and the indirectly assessed parts of the task were integrated with each other.
- Some of the practical work that students were required to carry out seemed quite lengthy.
- The instructions given for carrying out the practical work within each task were detailed.
- Overall, it was considered that the Chemistry tasks were more challenging than the Physics and the Biology tasks. Almost all of the Chemistry tasks took the full available time to complete.

Physics tasks

- Each task was based on an experiment. The title of the experiment was stated in a box at the beginning of the task.
- Students were allowed to select the method they used for the practical work and as a result questions had to be quite open.
- Students were required to select and set up the apparatus for the experiment.
- Students were told at set points during the practical work to alert the examiner so that the examiner could see what the student had done.

- A statement of theory related to the task was included in three of the tasks.
- It was reported during and after the trial that examiners had difficulty in distinguishing between two points on the *practical skills* marking scheme. These were "using the apparatus" and "making measurements or observations". In the case of most, if not all, Physics tasks this confusion was understandable as in all cases "using the apparatus" was in order to "make measurements".

6.11.4 Difficulty of tasks

A consideration with regard to judging the difficulty of tasks being set is that the practical examination in each subject is be set at a common level.

Biology tasks

Overall, the Biology tasks seemed to have been quite straightforward in general with a few being quite easy. This was emphasised by the level of detail in the instructions for students in some of the tasks.

Chemistry tasks

The Chemistry tasks were in general quite challenging and lengthy to carry out. Students were given detailed instructions.

Physics tasks

In the case of the Physics tasks, instructions were short and just stated what the student had to do. This was because of the range of different methods available to carry out each of the Physics experiments.

6.11.5 Key issues affecting task design – implications for new specification

Biology tasks

Many of the student practical activities in Biology are carried out over an extended period and could not be assessed during the practical examination. Other practical activities in the Biology specification involve fieldwork that cannot be assessed in a practical examination either. (see section 6.10 below.)

The overall lack of depth required in some of the specified practical activities in Biology makes it very difficult to set tasks with an appropriate level of challenge.

Chemistry tasks

The trial practical examination in Chemistry was very demanding of schools in that it required Chemistry teachers to prepare a very large number of solutions and other materials. When the new specification is being re-examined prior to implementation, this will need to be addressed, with particular reference to what stock solutions should be expected to be available as part of routine preparation for delivery of the course.

Physics tasks

In the case of the Physics tasks, instructions were short and just stated in a broad sense what the student had to do, thus allowing students the scope to choose the relevant and available equipment to complete the task. This approach was used because of the range of different methods available to carry out each of the Physics experiments.

6.11.6 Some suggestions arising from these reflections

The fact that there were characteristics of the tasks that varied between subjects means that a lot has been learned. As well as having implications for any future practical examination, this may have implications for the structure and content of the new specifications.

- In order to make it easier to adjust and align the standard, course coverage, and structure of tasks within and between subjects, it may be of benefit if all tasks in each subject were made up of two subtasks.
- The new specifications for the three subjects should be analysed to ensure that they facilitate the design of tasks for a practical examination that can achieve their intended purpose in the context of the constraints that apply. If this is not possible, then an alternative to this model of practical assessment should be considered.
- Consideration should be given, in all three subjects, to including instructions to students to alert examiners at particular points during their work.
- It may be worth considering the specification of a particular method for some of the student experiments in the new specifications, particularly with respect to Physics.
- Sub-tasks that use practical skills developed from undertaking the new specification and based on but not restricted to one or more or part of a specified practical activity could be used in the practical examination. These sub-tasks need not be restricted to specified practical activities that the students would have rehearsed, provided they are based on theory that is in the subject specification and detailed instructions are given in the sub-task. Nevertheless, the implications of the need for fairness between candidates as regards their degree of unfamiliarity with what they are being asked to do must be recognised. (See Chapter 9, Section 9.2.)

6.12 Content validity of the range of Biology tasks

It was noted in 6.9.5 above that many of the student practical activities in Biology are carried out over an extended period and could not be assessed during the practical examination. Other practical activities in the Biology specification involve fieldwork that cannot be assessed in a practical examination either. While this issue was known about right from the start, the practicalities of designing and implementing a suite of tasks for use drew it into sharp relief.

To quantify the extent of this difficulty, the EAMs for all three subjects were asked to review all of the practical activities in the new subject specifications and to identify in the case of each whether competence in executing that activity could be fully tested in a 90-minute laboratory-based practical examination of the type being trialled[13] (fully or partially). While it was judged that all of the activities in the case of Physics and Chemistry could be fully addressed, this was true for only two-thirds of the Biology activities, (10 out of 15) with a further two for which elements of the activity could be tested. In addition to the activities not testable in this way,

[13] It is noted that the practical assessment tasks are not necessarily exact replications of these activities, but this review nevertheless highlights they degree to which adequate coverage of the full target domain for the assessment can be achieved.

difficulties have been identified in relation to one of the 10 testable activities – dissection and display of a mammalian heart.

The fact that such a large part of the intended test domain is beyond the feasible scope of the range of possible tasks offered by this type of assessment is a worrying deficiency. It is best characterised as a deficit in the 'content validity' of the assessment or, in the language often used to describe the two main types of threat to validity, it is a form of 'construct under-representation'. This partial mismatch between the range of activities laid down in the Biology specification and those testable by the proposed assessment model, combined with the fact that the very large candidature for Biology makes it the most logistically challenging and costly of the three subjects to implement at full scale, mean that an alternative assessment model warrants consideration, as indeed has already been raised in consultative discussions. A coursework model, while not offering quite the same set of benefits as a laboratory based practical examination, brings a distinct range of opportunities instead, and may prove to be a more appropriate and feasible model for this subject even if not for the other two.

Chapter 7 The digital trial

The digital trial of practical assessment was intended as a small-scale trial of digitally accessed and reported practical examinations in Biology, Chemistry and Physics. Its principal aim was to serve as a proof of concept of using a digital format to present the examination and record and report students' practical activities. That is, can students be assessed on their practical skills using a digital assessment application that is cloud-based on a secure server in a secure digital environment?

The digital trial was similar in most respects to the main trial except that just three schools and 143 students were involved. The three subjects were examined in each school. The other principal points of difference arose from the fact that it was intended that as well as students accessing their trial practical examination tasks in the cloud, the examiners would access students' work online for marking. The tasks would be accessed online by the students using a laptop, iPad or similar interface and examiners would access students' completed tasks for marking in the same way. The three schools were Loreto, Navan; Villiers School, Limerick; and Luttrellstown Community School.

This chapter on the digital trial of practical assessment deals with the parts of the digital trial that were different from the main trial. These included students accessing their tasks online and answering the indirectly assessed part of the tasks online. The direct assessment of the students by the examiners and the actual examining and marking of students' indirectly assessed written work were similar to what happened in the main trial. For this reason, they are referred to only briefly in this chapter of the report. It was not intended that students in the digital trial would have paper task booklets and examiners would not have paper scripts to mark; both were to be accessed online.

7.1 Purpose of trial

According to the NCCA proposal for the trial, which was submitted in May 2017, the purpose of the trial was to gather information, among other things, on

- assessment items
- resources needed
- role of the examiner; technical competency implications
- role of the science teacher on-call
- the logistics of running the examination on a cloud-based platform
- suitability of the hosting environment
- security of the application
- database structure and integrity
- logistics of information movement from candidate to examiner.

7.2 Timeline of trial

The trial took place between 23 October and 25 October in three schools. It was intended that the digital trial would follow the completion of the main trial. However, as the main trial was delayed in many schools by the storm of 16 October and the resulting school closures, both trials ended up taking place simultaneously.

Two examiners were appointed by the SEC for each subject in the digital trial at the same time as examiners were being appointed for the main trial. As the number of examiners in the trial was so small, the role of advising examiner in the digital trial was undertaken by the three EAMs who were at the same time chief examiners for their respective subjects in the main trial.

Examiners were trained for the digital trial on 6 and 7 October along with the examiners for the main trial. They received a further day's training on Saturday 14 October that was directed at the technical aspects of the trial and at presenting the modified tasks for the digital trial. The digital tasks were adaptations of a small number of the main trial tasks for each subject. The adaptations were necessitated by the fact that the students would be accessing and submitting the tasks online.

Following the completion of the trial in the schools, the examination conference was held on Saturday 11 November and marking of the indirectly assessed part of students' tasks took place up to the end of November.

There was considerable interaction with the three schools by the NCCA in advance of the trial to familiarise schools and students with the technology involved. This interaction was supported by communications from the SEC to inform schools about the detail of the trial. It was intended that the tasks for the digital trial would be prepared by the NCCA and would have a significantly different emphasis to those in the main trial, which were based on the current syllabuses. It was intended that the tasks would require students to apply their knowledge in contexts that were unfamiliar. In the event, this was not possible, as schools were reluctant to have tasks in the trial examination that were not fully based on the existing syllabuses. As a result, the tasks that were used in the digital trial were adapted by the NCCA from tasks prepared by SEC for the main trial.

7.3 Technical aspects of the digital trial

This part of Chapter 7 is based on a draft NCCA report: *Report on the use of digital pro-forma in the trial of practical assessment in Leaving Certificate science* that was prepared in January 2018. The term pro-forma is used in the report to refer to the task booklets that were received by each student on her or his laptop or other device.

The draft NCCA report makes the point that while the purpose of the main trial was to examine the feasibility of the proposed model of practical assessment in terms of the logistics and cost, the purpose of the digital trial was to provide information on the possibilities of using technology to deliver the assessment. The report also points out that as the assessment was restricted to the current syllabuses, rather than on the practical activities in the revised draft specifications, the tasks were confined to a limited range of practical skills and the advantages of using technology in a practical assessment setting was not fully exploited. It

states that as a result, the NCCA report on the trial could only provide a limited amount of information about the potential role of technology in science practical assessment. The report states that the software available to students as they completed their tasks online included graphical analysis software.

The NCCA report states that the three schools for the digital trial were chosen based on their use of technology during class and their capacity to provide the appropriate infrastructure. One teacher in each school was the point of contact for the NCCA. The proposed trial was explained to them. In preparation for the trial, sample tasks were uploaded to the digital trial online computer application and a digital account was created for each student. The students were given sample tasks in advance so that they could practice using the assessment software and students made use of this access to practice on the sample tasks. As students continued to have this access while the trial was going on, this placed a strain on the online system due to an insufficient number of licences to accommodate the numbers going online. This was quickly resolved but not before one digital trial session was not able to proceed.

The report states that on the day of the trial students logged in to their digital account using their own login name and password. They were each given access to their examination task for the duration of the examination only. This was done online by the NCCA and examiners were informed.

In preparation for the marking of students' work following the digital trial practical examination, each examiner had been given a unique online account. Following the examination, the examiners would be allocated the tasks completed by a particular group of students as selected by the SEC.

A freely available app called *Logger Pro* was incorporated in the software used for the digital trial. *Logger Pro* is a graphical analysis tool used in some Irish schools for data logging. The *Logger Pro* used in the trial was stripped of some functionality and was deployed to Azure in a XenApp/Citrix application. The tasks were written in the app by the NCCA and a server image was account in Azure. This was then published, using the Citrix cloud environment, using Azure as the resource.

Because the application was deployed to the cloud, students could access the practical examination from any device. The only requirement of the school was that each student had to have the Citrix reader app installed on her/his device. One school used Chromebooks, in the second school, the students used their own iPads, and in the third school, 7-year-old iPads were used.

7.4 Outcomes of the digital trial

There are three sources of information on the digital trial used in this chapter. These are: the draft report furnished by NCCA as referred to in the previous section, the questionnaire responses received from schools and from examiners, and reports from implementation group meetings and meetings of the EAMs with the trial manager that took place during and after the digital trial.

7.4.1 NCCA draft report of January 2018

The trial practical examination sessions took place in the three schools as follows

- School 1: two examination sessions in Biology and Chemistry ran well. One of the two sessions in Physics did not run, as there were insufficient licences, as explained above.
- School 2: Two sessions in each subject ran well.
- School 3: The school's seven-year old iPads had not been updated and were not working properly. They had not been tested nor had students been given access to the sample tasks to practice on.

As the tasks used in the digital trial had been prepared for a written medium (pen and paper), they were considered less suitable for the students accessing them digitally. Some of the digital tasks had been slightly adapted by NCCA so that students could use the graphing software that was part of the digital examination mode. These adaptations were limited to requiring students to generate graphs and, in the case of Physics, to determine the slope of a line.

A disadvantage of using the digital medium that was highlighted during the trial was that students requested paper to do paper-and-pencil manual calculations. In many cases however, students who had experience of using digital media performed the calculations digitally. The ability of students to perform some of the calculations required in the digital trial could be measured in a written examination.

If practical assessment were to be incorporated in a digitally delivered form in the Leaving Certificate examination with the new subject specifications, the skills of writing scientific notation and of using software to complete calculations would need to be practiced by students.

In the main trial, students spent more time drawing graphs than in analysing them. The students in the digital trial did not spend time drawing graphs, rather they inserted the data they had collected into tables and their graph was automatically generated using this data. Students had the facility in the software to annotate the graph, change the axes and the scaling and perform multiple graphical analysis operations on the data using the software provided they were familiar with the software.

In the digital tasks, routine calculation such as calculating inverses could be done automatically within the data tables. For example, in one of the Physics tasks, students were asked to investigate the relationship between volume and pressure of a gas. They needed to graph the pressure against the inverse of the volume. Rather than calculating each inverse value, students were able to add a calculated column for 1/V and the data was generated automatically. To allow for the fact that students had not used this graphing software before, simple instructions were provided. In some cases, students followed these instructions correctly. Once the graph was generated, students used the software analysis tools to analyse the graph. Some students did this correctly; other students used manual calculations on paper instead of using the analysis tools.

7.4.2 The experiences of schools and examiners involved in the digital trial

Because of the small number of schools and students involved, the quantity of data available from the questionnaires of schools (principals, teachers, and students) is small. The questionnaire data available from the examiners is also of a small quantity. Schools and the examiners were requested to complete questionnaires that were the same as those used in

the main trial. Pressures of time precluded the preparation of specific questionnaires for the digital trial. Because the assessment objectives of the digital trial were based on it being a proof of concept, as outlined at the beginning of the chapter, the only data included in this part of the chapter is that related to the digital aspects of the digital trial. The experiences of the teachers, students, principals and examiners in relation to other aspects of the trial practical examination are similar to those reported for the main trial in Chapter 4.

7.4.2.1 The digital trial schools

One of the features that distinguished the digital trial from the main trial was the fact that in the digital trial the number of tasks used was small, three for Biology and three for Chemistry and four for Physics. Each practical examination session included up to twelve students. This meant that schools were required to have multiple sets of apparatus for each task. According to one teacher in a digital trial school, this meant that in some case insufficient equipment was available. A cause of class disruption referred to in by another teacher in preparing for the trial was the time needed by DSTs to organise students' computer equipment. When asked to comment on the fairness of the digital practical examination, one DST referred to difficulties with the software used in the trial. Ensuring that the technology for the trial was ready was stated by another DST to be one of the most difficult parts of their role. Principals of the digital trial schools did not refer in their comments to the digital aspects of the trial.

The questionnaires completed by 45 of the 143 students who took part in the digital trial were reviewed – in particular, the students' responses when asked what they found easiest about the trial, and what they found hardest. Also reviewed were their answers when asked whether they had any other comments about the trial.

When asked what they found easiest about the task, the responses of all the students except one were similar to those of the students in the main trial. One student referred to use of the iPad in the trial as making data recording easier.

When asked what they found hardest about their task, 13 of the 34 responses referred to problems with the technology and the remaining responses were similar to those in the main trial. Three of the responses are given below.

> The laptop. I found it hard to remember to use the laptop as I went along. The format was very unfamiliar. At one point, the laptop went to sleep and I lost the work I had done.

> Calculations. I couldn't properly type in my results and calculations in the computer.

> Using the app. It would not let me put in my data.

When other comments on the trial were invited, 19 students responded and 7 of them commented on the digital aspect of the trial and were critical of it.

> Wasn't fond of the paper online.

> There were some technical difficulties on the day of the trial.

7.4.2.1 The examiners for the digital trial

As with the analysis of the schools' experiences of the digital trial, the analysis of examiners' experience of the trial is based on the digital aspects only. As with the number of teachers, the number of examiners involved in the trial was low, at just six. What follows is based on the questionnaires completed by the examiners, and the centre reports that were completed by the Chemistry examiners and by one advising examiner.

It is evident that while there were difficulties with the technology in the digital trial in some schools, it worked quite well in others. One examiner reported that it was at the beginning of the examination, while students were setting up their tasks, that issues with technology were discovered and addressed. There was the situation described by a Biology examiner, who commented on the setting up of the students' computer equipment for the Biology trial

> We also had very slow iPads in the school and the students found the Citrix app very difficult to manage , they had to repeatedly log in and out and I wouldn't have known to do this only that Anna Walsh was there also the first day and she made sure it was working for that session. I learnt from this how to correct most of the problems that occurred digitally in the second session, and didn't need to go to paper version.

On the other hand, a Chemistry examiner reported

> The students had all practiced using the software and had pre-printed their login details and passwords.

There were difficulties during the digital examination such as that described by this examiner

> In school 1 the digital technology all stopped working. If this is to be done digitally, a minimum standard of technology must be had in all schools.

One Physics examiner mentioned that the school's technology expert was very helpful during the digital trial. The experiences of examiners in supervising the digital examination were different from the normal paper-based examination.

> As students were recording their answers digitally, I could not see at a glance their whole script. Almost all students asked for rough work paper.

While examiners reported difficulties due to the technology, this did not diminish their enthusiasm for practical assessment. The following comments by one examiner reflect this.

> Thoroughly enjoyed my time. Would love to be involved again in the future. Thank you for making me part of the team.

> Would love to see this practical element to the exams implemented. Do see huge resistance on the part of teachers due to the amount of preparation required by them for this process to run smoothly. I would hate to see a situation where students' results suffered due to the incorrect or insufficient preparation.

Examiner centre reports were completed by the two Chemistry examiners and by one of the advising examiners in respect of the digital trial. The purpose of the centre reports was to provide information on the experience of the examiner in the school during the trial.

The centre report of one examiner confirmed the situation reported earlier in one of the schools in the digital trial where the DST had little knowledge of the trial and students had not tried out their iPads before the trial. This was not the case in the other two schools. However,

in the case of another school, reference was made to logging in problems with some computers. Both examiners emphasised the need for students to be familiar with the technology in advance of a digitally based examination.

> Students need to have spent a lot of time practicing with the technology before they do the exam so they feel comfortable with it.

The examiner commented on their experience in the school where there had been little preparation for the trial.

> All the organisation seemed to be left to one member of staff. Disaster. All the iPads crashed and all the students ended up on paper. Students preferred the paper to iPads. They found it more user friendly.

While in another school, the situation was different.

> Very well. The Chemistry DST had spent a lot of time working with the students and using Chrome book and the software. 24 students could do the trial digitally. Again, this was down to the effort the DST had spent with the students in preparation.

Another examiner commented on the training received by the examiners on October 14 in NCCA Port Laoise.

> Training/practice session in Portlaoise insufficient. Not enough licenses available for all to log on. … Access to tasks in the cloud after Portlaoise training and before D2 trial was helpful to allow some practice.

It is evident from the examiners' responses to the questionnaires that the digital trial of practical assessment was insufficiently well prepared in some respects. In the case of one of the schools, students and the DST had not been sufficiently well informed nor had the devices to be used been sufficiently tested. The training of the examiners was deficient in that they were unable to access all of the tasks at their training session. However, this deficit was alleviated in the week before the trial.

Despite the difficulties, examiners reported that students succeeded in most examination sessions in access the tasks online and in saving them online. The fact that many students required paper to do rough work detracted somewhat from the fully online nature of their examination experience.

7.4.3 Reports from meetings

During and after the digital trial, meetings took place of the trial implementation group at which the digital trial was among the topics discussed. There were also meetings with the trial manager of the EAMs who had acted as advising examiners for the digital trial. The following are summaries of from the reports of these meetings.

The digital trial as it took place in each of the schools

The following summarises the trial in the three schools.

School 1
The digital examination for Chemistry and for Physics functioned satisfactorily. In the case of Biology, about half the students were able to use the digital mode satisfactorily.

School 2
Some of the students had successful experiences with doing the examination digitally.

School 3

There were considerable problems with regard to Physics and to Chemistry. The Biology examination was carried out digitally there. Some of the digitally produced Physics student work from this school was not available online for marking. The difficulties arose from the fact that outdated iPads were the interface for students. Students had not been prepared for the examination.

Overall, a considerable number of difficulties arose during the digital trial and as a result, students were frequently unable to complete the tasks entirely digitally or sometimes not at all digitally. Because of the difficulties with regard to students' accessing the tasks online, paper tasks had to be used in several cases. Many students used paper, either totally or partly or for rough work. This had implications for the marking process.

Difficulties/shortcomings of the digital software

During the trial, some shortcomings with regard to the software in use to deliver the tasks electronically to students' laptops or other devices became evident. These were principally:

- The text boxes in the tasks had characteristics that varied from text box to text box and therefore had no consistency.
- Students were unable to write chemical formulae and use other symbols, superscripts, etc that are part of science communication.
- As the students doing the digital trial used graphing software, the examining of their graphical work was quite different from that of the students in the main trial.
- Students were not able to type out calculations as part of the digital examination.

Marking of the digital trial

Because of the range of issues that arose with regard to the digital trial, marking students' work was challenging.

In the case of a small number of students, their digitally inputted work could not be accessed by examiners and so could not be marked. Even though some of the tasks used in the digital trial were the same as for the main trial, the changes made in the tasks to take advantage of the digital medium meant that some were quite different. In some cases, students in the digital trial had to use the main trial tasks as the digital element of the trial did not work. In other cases, where equipment failed during the trial, students answered from both the digitally adapted main trial tasks and the main trial tasks.

Because of all these issues, it was decided not to have examiners access the digital tasks online for marking. In any case, it should also be noted that online marking systems for recording marks and resulting had not been developed as part of the digital trial. It had been intended that examiners would view students' work on a screen and use paper to record their marks. The completed tasks of all of the students were printed out and examiners marked them as they had marked the main trial tasks. This was complicated by the fact that examiners in some cases were marking three versions of the same task. These were the digital version printed out, the paper version, and a combination of both in some cases. The examiners concerned showed adaptability and flexibility of the highest order in doing this work.

Conclusions from the discussions on possible future implementation of digitally based practical assessment

It is accepted that as a proof of concept the digital trial enjoyed some limited success. In some cases, students were able to access their tasks online, report on them, and answer the questions in the task fully online without any requirement for pen and paper. However, most students were unable to perform the tasks without needing paper for calculations, and therefore the tasks were not independent of paper and therefore were not fully digital.

As at least one school in the trial did not have the digital capacity to run the digital trial, and this is the likely situation in many schools, this type of examination should be deferred until schools have sufficient digital capacity. There is therefore, at present, no question of the full implementation of digital tasks due to the excessive potential for systems failures.

An outcome from the digital trial that could have consequences for an implementation of practical assessment in the Leaving Certificate examination is that carrying out multiples of the same task was reported to have been successful during the digital trial in some schools.

7.5 Conclusions with regard to the digital trial

The digital trial set out to test two concepts. It was intended as a test of the feasibility of students accessing tasks online, completing them, and submitting the completed tasks online. It was also a test of examiners accessing students' work online and marking that work.

The digital trial was a limited success on one of these counts, that of students accessing and submitting their examination materials online. It was limited because it did not work for all students, and successful because many students succeeded in doing this. It was not possible to test the online marking of students' tasks. The software provided had shortcomings, some of which are referred to above.

While it is clear from the trial that online examinations of the type trialled are possible, any system-wide adoption of them would need to be deferred until further developments take place. Principal among these are

- Development of more user-friendly software for the online examination
- Further acquisition by schools of suitable hardware
- Strengthening of the schools' IT networks
- Further acquisition of the necessary skills by school staff
- Further development of online-adapted assessment materials
- Development of SEC IT resources.

Chapter 8 Rolling out the model nationally – scalability, cost, and scheduling

This chapter deals with the scalability and cost of the proposed model of practical assessment, including consideration of feasible ways to schedule such an examination at scale.

One of the objectives identified for the trial was that it should give good information on the impact of the trial on the schools involved. The availability of sufficient teachers who would be available to act as visiting examiners is a key issue in deciding on a roll out of practical assessment. It was important that the trial should contribute useful information on this in particular. The possibility of differentiating between the three subjects in the implementation of practical work assessment was also identified as worthy of consideration. As well as giving information on the trial and the findings arising from its evaluation, the final report is to include cost estimates for a full roll-out by SEC of the arrangements and identify other issues for consideration in advance of such a roll-out.

Some of the findings from the trial that have a particular consequence for full implementation are addressed below, along with other issues outside the scope of the trial but which affect the feasibility of a roll-out.

8.1 Adequacy of laboratory facilities

It was beyond the scope of the trial to carry out a comprehensive assessment of the adequacy of science laboratory facilities across the country. While the schools involved in the trial were found, for the most part, to have facilities of a sufficient standard to allow the assessment to take place, there were nonetheless difficulties in a number of schools in this respect that made the experience less than ideal. Furthermore, it must be remembered that these schools were selected from a pool of schools that volunteered to participate, and therefore cannot be considered representative. Schools where the facilities are inadequate to support high-quality practical work would have been unlikely to apply to participate.

Accordingly, any evaluation of the current laboratory facilities in second-level schools remains to be carried out. On the one hand, it could be argued that the facilities, equipment and materials required to carry out this assessment are no more than those required to deliver the existing curriculum in the intended manner, so that schools should already have them in place. However, this seems at best a risky assumption to make. In the absence of more recent comprehensive information, it may be noted that the *Report and Recommendations of the Task Force on the Physical Sciences* (DES, 2002) estimated, based on 2002 costs, that providing access to 'first-class laboratories and equipment in all post-primary schools' would entail a capital investment of €142.8 million over two years. The extent to which the situation has changed since then remains unclear. Furthermore, what the task force regarded as 'first-class laboratories and equipment' is not necessarily the same thing as sufficiently good laboratories and equipment with which to carry out a practical assessment.

8.2 Science teacher from school to be on call

Among the issues that it was hoped would be answered by the trial was whether the DST would need to be a teacher of the specific science subject being assessed or whether another science teacher might do, and, related to this question, whether one DST could provide on-call assistance to all of the practical science tests in the school on a given day. The experience of the trial was that the DST must be a teacher of the subject being assessed. This was due to the nature of on-call assistance required, as, to be of sufficient practical assistance, the DST needed to be familiar enough with the laboratory facility, the equipment and materials, and the demands of the subject. (The question of whether one DST could provide support simultaneously across a number of assessment sessions in a school on the same day therefore becomes largely moot.) In schools with a laboratory technician, the laboratory technician was able to provide the required support across all three subjects. In such cases, it may be possible to support the examination without the presence of the teacher of the specific subject, through the support instead of the laboratory technician and a non-subject-specific member of the teaching staff. In the more general situation (where laboratory technicians are not available,) the level of overhead on science teachers in schools is of significance in planning for full implementation. As a consequence of the need for the DST to be a teacher of the subject being assessed, the scheduling of the practical examinations will need to be such that they are not taking place in particular schools at the same time as those schools have released teachers to act as external examiners.

8.3 Standard equipment and supplies lists

In advance of the conduct of the trial, the schools involved were provided with a list of the materials and equipment required. Schools in general had sufficient chemicals and other supplies, although there was variation between the schools as to whether they had, or had in sufficient quantities, the materials and equipment required. In addition, there was a considerable overhead on science teachers in advance of the trial in making ready the equipment and materials for the trial. This was particularly acute in the case of Chemistry, due to the demand to prepare chemical solutions and compounds. Payment for preparatory work by the SEC to teachers is a feature of the practical tests in certain subjects (Junior Cycle Home Economics and Metalwork; Leaving Certificate Construction Studies and Engineering). However, this payment is not a standard feature in all subjects with an externally examined second component. The experience in the trial would suggest that the additional work required of teachers of these subjects in undertaking the advance preparation demanded will need to be addressed, whether through financial or other compensation, provision of dedicated substitutable time, or other additional support. It is also essential that there is some way of quality controlling the preparatory work – for instance, to check that solutions for the Chemistry practical assessment are prepared to specification. It is therefore likely that, not only will there be costs associated with having this preparatory work completed, but also some associated with examiners checking equipment and preparations in advance of sessions beginning.

8.4 Time of year for the practical examinations

A number of options are proposed for consideration:

1. Schedule the tests to run concurrently with the oral and practical tests shortly before or after Easter
2. Using a model similar to the current oral and practical model, schedule the tests at another time during the school year at which schools may be more willing to bear the absence of their science teachers
3. Schedule the tests during the February mid-term break or during the Easter Holidays
4. Under a model similar to PDST or JCT, establish a temporary team each year to deliver the tests over a longer period of time.

These options are considered in turn below.

1. Schedule the tests to run concurrently with the orals and practical tests shortly before or after Easter

The current oral examination model operates as follows. All oral examinations are conducted by visiting examiners over a two-week period, either before or after the Easter break depending on how early or late Easter falls. In 2018, the oral tests were conducted after Easter in the period of 10 working days from the Monday 9 to Friday 20 April. Although the oral examinations are run over a 10-day period, the current model is actually a 2 X 5 Day Model. The 10-day period is spilt into two units of five days each. The country is divided into two regions for logistical purposes:

- South Country and North Dublin (Region A)
- North Country and South Dublin (Region B).

Schools in Region A have Irish examinations for the first week while the modern language examinations are being conducted in Region B. The situation is reversed in Week 2. Examiners are normally appointed for a five-day examining period. In effect this means that examiners are appointed to either Region A or Region B for a five-day period.

Table 8.1 Current scheduling model for oral examinations

	Day 1–5	Day 6–10
Region A	Irish	Modern European Languages, Japanese
Region B	Modern European Languages, Japanese	Irish

For the SEC, a very significant problem with the current model is the acute difficulty in recruiting sufficient numbers of examiners for what is an essential function. The recruitment problem can be attributed in part to the understandable reluctance by many schools to release staff at such a critical time in the school year for a number of reasons:

- Unwilling to bear the loss of five days tuition time and the consequences for the completion of syllabi at a critical time in the academic calendar.

- Creates significant difficulties for schools in arranging substitution cover for those absent on examining duties. With the heavy demand for qualified substitute teachers over the two-week examining period, schools have increasingly found it difficult to recruit qualified teachers.
- The shortage of qualified substitute teachers means that, in most cases, students in both examination and non-examination classes can only be supervised (as opposed to being taught) when their teachers are released for examining duties.

Of those schools that do release staff, many operate an annual quota system and only release a proportion of the teachers in a given subject.

Even if schools are prepared to release staff on examining duties, individual teachers may be unwilling to make themselves available. Reasons cited for this include:

- unwillingness to be absent from their own students during examinations.
- perceived lack of adequate remuneration for examining duties
- the administrative burden
- the length of the working day if travelling long distances
- family commitments and child care issues.

Within the current assessment regimen, the closest visiting-examiner model to that envisaged by the proposal at hand is that of Junior Certificate Home Economics. Examiners visit schools to undertake an individual assessment of candidates as they undertake practical cookery tasks. While the assessment is individual, the candidates are operating in the kitchen with other candidates, with typically 10 candidates assigned to a session. The conduct of these tests requires a team of 260 examiners to undertake tests with 23,000 candidates. Examiners who are serving Home Economics teachers will not be assigned for more than 5 days; retired and unemployed teachers can be assigned for the full period of the practical tests. Ideally, only serving or recently serving subject teachers, familiar with the demands of the specification, would be appointed to these roles. The reality is that the oral and practical tests would not be delivered without the involvement of retired teachers available to provide their services over the 10-day timeframe. Recent years have been characterised by a severe difficulty in recruiting the required numbers of examiners and filling the posts has required considerable time and effort by the SEC, involving repeated appeals to individual teachers, schools, and their collective representative organisations.

This year saw the SEC take the unprecedented step of extending the normal two-week timeframe for the conduct of the Junior Cycle Home Economics practical tests to a third week.

In light of ongoing examiner supply issues and impact on tuition time in schools, the current model of external examinations during term time, for the oral and practical tests, is increasingly regarded by the education stakeholders as unsustainable. Seeking to introduce a new practical component in the science subjects under this model, particularly in light of the findings from the trial around the need for science teachers to be in their own school while the tests are being conducted in their subject, could not be recommended.

Any new structure for conducting the oral and practical examinations must address the problems inherent in the current system and must provide a workable, practicable solution for the SEC, schools, examiners and candidates. For it to be successful, any new model would have to achieve the following outcomes:

- enable schools to release staff on examination duties with minimal disruption to schools in terms of encroachment on tuition time and teacher withdrawal from schools
- provide a sufficient supply of examiners
- be practicable for the SEC to operate
- allow a reasonable itinerary for each examiner
- be practicable for schools to operate
- safeguard the integrity of the examination system
- not impact negatively on candidates
- provide value for money.

2. Schedule the tests during the Easter Holidays or the February mid-term break

In previous engagements with the education partners about the timing of the oral and practical tests, some of the models that have been considered involve holding these tests outside of term time. One model proposed was to run these tests partly during term time and partly over the Easter holidays. The proposed change was broadly supported by all of the education partners involved in a working group (comprising school managements', unions' and parents' representatives) but could not be further progressed at that time due to the additional costs of school opening arrangements and school transport during the school holidays. In more recent engagements with school managements' and unions' representatives, more divergent views have been expressed as to whether the proposed solution would actually address the problem it was designed to solve – that is, would teachers be willing to forego their holidays in order to undertake external examinations duties and, if so, at what cost? Engagement between the SEC and the stakeholders on alternative models is continuing.

3. Using a model similar to the current oral and practical examination model, schedule the tests at another time during the school year at which schools may be more willing to bear the absence of their science teachers

It has been suggested that if the oral and practical tests were held during term time at a different time of the school year, this might enable greater ease for schools to be able to release teachers to external examining duties. A model similar to the current oral examination model, with tests arranged on a regional basis to facilitate teachers being available to their own students to the greatest possible extent, and also being available for examining duties with the SEC, might be arranged as in Table 8.2.

Table 8.2 Scheduling model for science practical examinations under option 3

	Day 1–5	Day 6–10
Region A	Biology	Chemistry, Physics
Region B	Physics, Chemistry	Biology

The suggested timing is the first fortnight in December of sixth year. The timing needs to allow sufficient time for students to have covered the practical element of the course of study. Timing these tests in January or February is unlikely to find favour at school level as it coincides with the preparation and holding of the mock examinations. Any later than February is likely to result in the same issues as the SEC is currently faced with in the current model of delivery of oral and practical tests.

Issues for consideration include:

- the amount of practical work that students will have completed by that time – could they be reasonably expected to have undertaken all of the practical work covered in the specification?
- the possibility that practical work would be taught separately from the underlying theory in preparation for the December practical assessment
- whether the timing would allow more teachers to be released from schools and/or to make themselves available
- availability of qualified substitutes.

4. Under a model similar to PDST or JCT, establish a temporary team each year to deliver the tests.

Consideration should be given to the annual establishment of temporary team of external examiners to deliver the practical tests in schools. The model would see teachers of Biology, Chemistry and Physics being seconded each year to the SEC for a three-month period for the purposes of visiting schools to conduct the tests. A secondment model will mean that schools should be able to put in place a temporary contract for a qualified substitute for a three-month period rather than seeking *ad hoc* arrangements for qualified substitutes for five days. Issues for consideration include:

- the timing – optimum time of year for the examining period to commence and end taking account of periods of school closure; mock examinations and proximity to the written examinations.
- avoiding issues that candidates will be advantaged/disadvantaged as a result of the timing of tests in their schools over the examining period
- assessing whether establishing a temporary team is sustainable over time and the measures that might be put in place to achieve this
- the number of examiners required to deliver the tests in the timeframe
- scheduling of tests in schools
- payment model, which will be salaries plus travel and subsistence.

It is not clear whether a model that requires secondment of a smaller team of examiners over a three-month period would be more attractive or less attractive to potential examiners, but we consider it likely to be more so. Likewise, it is not obvious to see whether it is easier or harder for school management authorities to facilitate release of smaller numbers for a longer period than the release of larger numbers of examiners for one to two weeks, but it seems likely that it would be easier to find a suitably qualified substitute for an extended contract than a short one, and of course far fewer such substitutes would be required. On the other hand, in many cases, school authorities are currently releasing examiners without having access to a substitute qualified to teach the subject concerned, and they would presumably not be prepared to do this for a much longer contract period.

Any secondment for a period less than an academic year is likely to be at least to some degree problematic for school management authorities. For this reason, consideration might also be given to looking at options that combine to address the SEC's examiner needs with other system needs. For instance, a model might be considered that would involve seconding, on a longer term basis, a team of science teachers who might deliver ICD for certain parts of the year and work as SEC practical examiners for other parts of the year.

8.5 Examiner requirements

Using data on schools from the Department of Education and Skills statistical report of June 2016 and SEC statistics of the 2016 Leaving Certificate, Tables 8.3 and 8.4 extrapolate demand for examiners under a five-day and a forty-day examining model.

The first table below assumes implementation as trialled. The experience of the trial and other experience of organising practical tests suggests that, if a maximum of 12 candidates per session is allowed then the average number of candidates per session is likely to be approximately 8. Likewise, if the maximum number of sessions per day is 3, (with no more than 2 on the first day in a school) the average number of sessions per day will be approximately 2. These averages are used in the first table below.

However, the experience of the trial suggests that examining up to 12 candidates simultaneously in a session is a challenge, and that 8 might be more reasonable for roll-out. This would yield an average across sessions of approximately 5 candidates per session, and this is used in the second table below. It is considered that aiming for three sessions a day, while challenging for all involved, is feasible. Furthermore, reducing this to two would extend the number of days in a given school, thereby increasing the level of disruption to the normal work of the school.

Table 8.3 Arrangement A – a maximum of 3 sessions per day and a maximum of 12 candidates per session

Subject	Schools	Candidates	Average number of Candidates per school	Average 2 per day and 8 per session		Five-day examiner requirement	Forty-day examiner requirement
				Sessions	Lab Days		
Biology	687	34101	49.9	4250	2100	420	53
Chemistry	581	9089	15.6	1125	600	120	15
Physics	537	7753	14.4	1000	500	100	13

Table 8.4 Arrangement B – a maximum of 3 sessions per day and a maximum of 8 candidates per session

Subject	Schools	Candidates	Average number of Candidates per school	Average 2 per day and 5 per session		Five-day examiner requirement	Forty-day examiner requirement
				Sessions	Lab Days		
Biology	687	34101	49.9	6800	3400	680	85
Chemistry	581	9089	15.6	1800	900	180	23
Physics	537	7753	14.4	1600	800	160	20

Under the current eternal examining model, examiners are paid a per-candidate fee plus travel and expenses. For the purposes of the trial, these costs were set at €10 per candidate for the practical skills assessment (examining on the day) and €4 per candidate for the assessment of results and analysis (marking the booklets). All SEC contract staff fees have increased by 1% for the 2018 examinations.

Cost of monitoring is estimated at 30% of the candidate fees. Cost of travel and subsistence is estimated at 135% of candidate fees. An average substitution per diem rate of €164.26 has been used.

The estimated annual costs of examining under full implementation of a five-day model, under each of the two arrangements A and B above, and excluding headquarter payroll costs, are as shown in table 8.5 below.

Table 8.5 Estimated main costs for arrangements A and B

	A	B
Biology		
Candidate Fees	€482,188	€482,188
First Day Fees	€86,650	€140,291
Monitoring and Advising Fee	€144,656	€144,656
Two Day Training Fee (substitution)	€137,978	€223,394
Travel & Subsistence Fees	€650,954	€650,594
Total	**€1,502,426**	**€1,641,123**
Chemistry		
Candidate Fees	€128,518	€128,518
First Day Fees	€24,757	€37,136
Monitoring and Advising Fee	€38,555	€38,555
Two Day Training Fee (substitution)	€39,422	€59,134
Travel & Subsistence Fees	€167,073	€167,073
Total	**€398,325**	**€430,416**
Physics		
Candidate Fees	€109,627	€109,627
First Day Fees	€20,631	€33,010
Monitoring and Advising Fee	€32,888	€32,888
Two Day Training Fee (substitution)	€32,852	€52,563
Travel & Subsistence Fees	€147,996	€147,996
Total	**€343,994**	**€376,084**
Grand total (all three subjects)	**€2,244,745**	**€2,447,623**

Additional costs will arise for SEC, such as fees for drafting and setting, cost of materials and stationery, and costs associated with printing and translation. These are unlikely to be significant, but it is unlikely that the management of these examinations could be serviced by the existing complement of EAMs for the science subjects, or its administration by the existing staffing complement in the *Orals and Practicals* section, so additional staff costs will also arise. Furthermore, the above cost estimates do not include any provision for payment to teachers for preparatory work or substitutable preparation time.

8.6 Numbers of teachers and examining personnel

Ireland's examination and assessment system relies on almost complete external assessment for all components of examinations in both Leaving and Junior Certificate programmes. The majority of personnel engaged in examining duties are drawn from the serving second-level teaching community.

To illustrate how the scale of the operation proposed would sit within the scale of existing arrangements across other subjects, total numbers of contract staff in various categories are shown in Table 8.6 below.

Table 8.6 Contract staff statistics 2013–2017

Year	Oral Examiners	Practical Examiners	Written Examiners	Superintendents	Candidates (LC and JC only)
2013	1,079	1,241	4,223	4,861	112,589
2014	1.122	1,238	4,362	4,975	114,353
2015	1,146	1,256	4,307	4,987	114,295
2016	1,153	1,277	4,419	5,119	115,955
2017	1,169	1,335	4,376	5,131	117,424

The numbers above include contract staff at all grades in each of the categories presented. The table excludes LCA; candidates and examiners.

Chapter 9 Towards a policy decision on roll-out

We return now to the stall that was set out in Chapter 1: what does the trial tell us about whether the proposed assessment model can and should be rolled out, and what factors need to be considered when making that roll-out decision.

In Section 1.4, we set out a range of questions that the trial would help answer. We revisit those questions here, grouping them under various headings and dealing with related considerations as we go along.

9.1 Validity of the tasks and the assessment process employed

The first question asked was: *can a suitable range of tasks be prepared for each subject, providing an appropriate vehicle through which the target skills can be validly and fairly assessed?*

The answer to this question is fundamental to evaluating the entire enterprise. Crucially, though, it cannot be answered without being clear about what exactly are the *target skills*? Since a clear specification of the attribute to be measured is the first step in good test design, we turn for an answer to the relevant documentation– the draft specifications and associated documents. The specifications do not include an explicit statement in exactly such terms as to the attribute(s) that the practical examination is intended to measure, but instead seek to make this clear by describing the characteristics of a high, moderate, or low level of achievement in the practical examination. All three subjects have similar such statements, and the one for Physics is cited here:

> A *high level of achievement* in this component is characterised by demonstration of a comprehensive range of manipulative techniques in experimental activities. Candidates make and record observations and measurements with a high level of accuracy and precision. In almost all cases candidates recognise and describe trends and patterns in data and use physics knowledge and understanding to account for inconsistencies and anomalies. Candidates accurately interpret and analyse experimentally derived data; manipulation of the data is almost flawless. In all cases candidates link theoretical concepts to interpretation of experimental evidence.
>
> *Leaving Certificate Physics – draft specification*

The characteristics of moderate and low levels of achievement are described similarly, with the degree of competence scaled back appropriately.

On the basis of these statements of what the assessment is intended to measure, the answer to the question as to whether the tasks can measure the right skills is, broadly, 'yes'. The range of tasks employed and the assessment process trialled did indeed test in full for the presence of all of the characteristics described above. It is therefore reasonable to infer from a high score in an assessment of this form that the candidate displays the above characteristics of high achievement.

Notwithstanding this, a view emerged from early on the project that the tasks were not adequately addressing the full range of skills that are important when engaging in practical experimental work in science and which are emphasised in modern science curricula. In

particular, there was much discussion as to the degree to which the candidates should be required to apply their practical and investigative skills to less familiar 'unseen' scenarios and contexts. The draft subject specifications envisage open exploratory work to be a key feature of the practical activities that students would engage with over the course of the programme of study. It was argued that an assessment of practical work should reflect this by presenting candidates with, at least to some degree, less familiar scenarios in which their existing science knowledge is insufficient to allow them to know in advance what they 'should' find.

Such a purpose for the assessment goes well beyond the statements in the subject specifications as to what the practical assessment is intended to measure. Nevertheless, it is not an unreasonable aim, and such an aspiration is reflected in some of the documentation developed by the NCCA executive in the early stages of the development of the specifications. The document *Senior Cycle Sciences – Assessment Outline*, developed later, remains close to the specifications in its description of the two proposed components, making clear that the practical examination is focused on 'practical competence and procedural understanding' and states that 'the main focus of the practical examination is on the use of equipment for collecting, recording and interpreting of data' (p.6). The capacity to deal with the unfamiliar is described as being part of the target domain of the written rather than the practical examination. On the other hand, an earlier document *Assessment of Leaving Certificate Science* includes the phrase: "…despite being unknown and unrehearsed, the assessment tasks…" (p.8) and later says "the short practical tasks are designed to provide an opportunity for learners to demonstrate the practical skills […] in unrehearsed contexts." (p.11).

This is a crucial difference in perspectives on what the practical examination should measure. If we confine ourselves to what the subject specifications say is its purpose, then the trial has shown us that we can indeed use it to measure what we wanted to measure. But the capacity to apply a range of practical and cognitive skills in unknown or unrehearsed contexts, as indicated in the earlier *Assessment of Senior Cycle Science* document, is an entirely different matter. These tasks have not been designed to do that. If the purpose is to be expanded to include this, then these kinds of tasks and this assessment process will not do the trick. The question naturally arises as to whether the tasks and/or the process could be adjusted so that the practical examination *does* measure this more ambitious attribute, and this is addressed in the next section. We might surmise that a recognition of these likely difficulties was one of the reasons why there is no reference to dealing with unrehearsed and unknown tasks when specifying the purpose of the practical assessment in the later documents.

Before we move on, though, we should deal with two more of the questions posed in Section 1.4 and which are part of the technical assessment of the test's validity. First, *can the candidates carry out these tasks (and complete the booklets) comfortably in the allocated time?* While some of the tasks may need to be somewhat shorter than those used in the trial, shortening them should not pose any great difficulty, so the answer is, broadly, 'yes'. Second, c*an examiners reliably assess the practical skills that they observe the candidates displaying?* The answer to this is also 'yes', as satisfactory levels of marking reliability were achieved in the trial. This assertion is subject to a caveat regarding the maximum number of candidates in a session, which is dealt with later.

9.2 Could the assessment do more?

To serve the more expansive purpose referred to above, the tasks would need to present candidates with situations with which they are less familiar, but where they can bring to bear their existing knowledge, skills and understanding to explore and investigate the situation. While this is theoretically possible, it runs into a number of difficulties in practice, largely arising from the nature of the Leaving Certificate examination as an externally assessed examination and its particular role in the education system.

9.2.1 The fairness imperative

In the context of a high-stakes examination like the Leaving Certificate examination, fairness for all candidates is an over-riding consideration. No candidate or group of candidates can be allowed to have an unfair advantage over any other arising from the nature of the test or the manner of its implementation. This has always been viewed as an imperative that trumps all other aspects of validity and has in the past imposed constraints on the structure and content of assessment for the Leaving Certificate. Some potential sources of unfairness can be addressed procedurally. For example, if potential unfairness arises from the fact that different candidates are assigned different tasks that cannot be guaranteed to have equal levels of difficulty, it is possible to use suitable statistical or other techniques to re-scale the results so as to address this problem, as mentioned in Section 5.5 of Chapter 5. Other threats to fairness are more difficult to deal with. In the context of considering designing sets of practical assessment tasks that require candidates to engage with unseen contexts, the main threat to fairness arises from the fact that – no matter what implementation arrangements are made – it will not be possible to have an externally assessed practical examination in which all candidates take the test at exactly the same time. At a minimum, the examinations will take place over a period of two weeks, involving at least thirty time-slots, and, as is clear from the discussion of roll-out options in Chapter 8, probably much longer.

When all candidates do not take a test at the same time, the question of the security of the test items arises. In this day and age, for any high-stakes test, it is impossible to ensure that details of an assessment taken on one day will not be available to some candidates – or indeed many of them – shortly thereafter. How then can we ensure that some of those who take the test on a later date do not have an advantage over those who take it on an earlier date? Several possible scenarios arise, and examples from the Leaving Certificate and other familiar testing programmes are offered below.

9.2.3 Fairness through advance knowledge for everybody

One possible scenario is a case in which any candidate having advance knowledge of the test materials is not considered a problem. Fairness can then be achieved by ensuring that *all* candidates have access to the material in advance. This arises, for example, in the case of the life-drawing examination in Leaving Certificate Art (in which candidates make drawings of a model in two different poses). For this test, candidates and teachers are informed about two weeks in advance of the two-week assessment window as to what poses the model will be in. It is possible that candidates may practice drawing models in these poses and it is also the case that candidates taking the test towards the end of the window will have more time to do this than those at the start. Nevertheless, it is considered that, since this examination is testing

drawing skills that are largely independent of the pose and are developed over considerable time, no significant differential benefit accrues.

This solution is only satisfactory in cases where advance knowledge of the precise details of the test does not compromise its integrity or validity. While advance knowledge of the task is not considered to compromise the validity of the life-drawing examination, the same cannot necessarily be said for another example from the Leaving Certificate examination: advance knowledge of the picture sequences in the oral language examinations[14]. Another scenario worth noting is that of the practical examinations in Engineering and Construction Studies. Since school facilities are typically not such as to allow all candidates to take the test at the same time, schools had, up until 2017, freedom to organise these examinations over several sessions over a specified period of two weeks. In this case, although the test is primarily a skills test akin to the life drawing examination in Art, it also tests the candidates' capacity to interpret and analyse drawings and to plan the execution of the practical work. Consequently, advance knowledge is considered to convey some advantage. Efforts were made to keep the test materials secure over the course of the available window, but this proved increasingly difficult as time went on. Accordingly, this arrangement has been discontinued. As of 2018, these tests are conducted over three consecutive days and a different version of the test is used each day. This example illustrates that even a comparatively modest level of potential advantage cannot be tolerated.

9.2.4 Fairness through advance knowledge for nobody

If the practical assessment tasks for science remain as trialled, without seeking to test candidates' engagement with unfamiliar contexts, then the examination can be considered to be like the life-drawing examination in Art, with its integrity uncompromised by advance knowledge of the tasks. Otherwise, it cannot. This brings us to another potential method for ensuring fairness when not all candidates take the test together: the use of different test forms each day. Indeed, this arrangement was suggested in one of the NCCA supporting documents:

> Each sitting of a practical examination in any one subject will be based on different tasks to any previous sitting, so that students completing the assessment on different days will not have an unfair advantage over those who completed the assessment earlier.
>
> *Leaving Certificate Science Subjects – Working towards Implementation, p.9*

Depending on the number of days on which the test occurs, this can necessitate the preparation of a large number of test forms annually, which may not be feasible. Indeed, it is hard to see how it would be feasible in the case of a science practical examination of the type trialled here, which already involves 12 distinct test forms for each session. Even with a two-

[14] At the design stage, the intention was that candidates would be presented with a sequence of pictures (a storyboard) that they would not have seen in advance. Oral language skills developed in a natural and appropriate way would allow them to describe and discuss what was shown. However, due to these examinations being carried out over a two-week period, many candidates taking their test towards the end of this window would already be aware of the sequences and could have practiced specifically for them. Since this unfairness between candidates could not be tolerated, the 'unseen' aspect was abandoned and the sequences were made available to all in advance. This solved the fairness issue, but it also changed what was being tested and had a negative washback effect by encouraging inappropriate rote-learning.

week model, this would require the preparation of 360 distinct task booklets in each subject annually, which is clearly unworkable.

9.2.5 Item banking

Another means of ensuring fairness for all candidates is by using a test design that involves each candidate engaging with a random selection of tasks from an item bank. If the item bank is large enough and is considered to cover the entire intended test domain, its security is not regarded as a concern. An example of this is the driver theory test. There is a bank of 1250 test questions and each candidate is given a random selection of 40 questions. The entire item bank is publicly available for purchase. This is not considered a threat to the integrity of the test, as the view of the test developers and their client (the Road Safety Authority) is that attempting to learn off the answers to the entire bank of questions is, in essence, equivalent to acquiring all of the knowledge that the test is designed to assess. There is perhaps some potential in exploring this model, (a large bank of tasks that would remain largely static over time,) but two interlinked difficulties remain: first, how large would an unsecured item bank of practical tasks have to be in order that we can still be reasonably confident that a candidate engaging with a randomly selected task from that bank is engaging with an unseen or unrehearsed scenario? Second, does the range of subject content in the specifications admit the creation of such a sufficiently large range of meaningfully distinct tasks? It seems implausible that we would find ourselves able to answer 'yes' to both of these questions.

9.2.6 Implications for engaging with 'unseen' material

It is in the context of all of this that we must consider what the proposed form of practical assessment in the sciences can hope to measure. It will not be possible to keep the test materials secure beyond the date of first use. So, the next question that must be asked is: does advance knowledge of the test materials (perhaps a week or two in advance) convey any advantage? On the one hand, given that there is a set of 12 tasks for each session and each candidate has no way of knowing which task they will be given, a few days' advance knowledge does not provide sufficient time to practice completing the tasks in order to develop any significant level of task-specific skill beyond the more generic practical skills already acquired. On the other hand, this argument only applies to the practical skills tested and not necessarily to certain forms of cognitive skill. If it is intended that the tasks would require candidates to do more than display understanding of and execute comparatively routine procedures and cover known territory, then advance knowledge could offer significant advantage. For example, if the tasks required candidates to apply a known procedure in a somewhat unfamiliar context and to draw conclusions that are not immediately obvious, then this is certainly a case where advance knowledge of the question asked is an advantage, as it allows the candidate to research and/or consult as to the answer.

This is a constraint that cannot be overcome in an assessment of the type proposed. We have to be fully satisfied that no significant advantage can be gained from advance knowledge of the suite of tasks being used in a given year. Given the inevitability that some candidates towards the end of the assessment window will be aware of some or all of the tasks used, the only way to ensure fairness is to make that knowledge available to all – in short, advance publication of the full suite. This can be achieved either through a large openly available item-bank of tasks that remains largely fixed over a number of years and from which any might be

drawn in any particular test session, or the annual preparation of a suite of 12 or more tasks, which are published shortly before the assessment window begins.

Irrespective of the details, the crucial point to note in the context of making a decision to roll out this model of practical assessment is that it will **not** be possible with this model to test the candidates' capacity to apply their knowledge or skills to unfamiliar contexts. Those cognitive skills will have to remain in the target domain of the written examination only. This is consistent with the purpose of the assessment as expressed in the subject specifications themselves, but not with the purposes expressed in some of the ancillary documentation.

9.3 Impact on teaching and learning

Among the questions listed in Section 1.4 is the following critical one: *What range of impacts is such an examination component likely to have on teaching and learning?* There is no doubt that teaching and learning practices in Irish schools are heavily influenced by the Leaving Certificate examination. Introducing the proposed model of practical assessment will certainly cause a change in behaviour. Some evidence for the changes that are likely to arise comes from the feedback from participants, but it is also possible to use experience and logic to anticipate what is likely. There may be both positive and negative impacts.

Likely positive impacts

It is certain that many teachers and students will pay more attention to carrying out practical work in science. They are more likely to pay explicit attention to developing the skills that they perceive the practical examination to be testing. Since there is a generally held view that practical work does not currently receive enough attention, this is positive. There is little doubt that students will end up doing more practical work and becoming more skilled at carrying out practical work if this form of assessment is rolled out.

It is worth noting the following finding mentioned in Section 6.3: *Many of the errors or deficiencies observed in the practical skills displayed by students in the trial are consistent with poor or limited exposure and training in practical laboratory skills as part of the current delivery of the subject.* This indicates that, even within this sample that might be expected to be skewed towards schools in which practical work is valued and routinely carried out, students are displaying deficiencies that suggest inadequate exposure to practical work. Thus, the extent of the impact that this component would have on practice in schools should not be underestimated.

A secondary positive impact that can arise through the introduction of such a mode of assessment arises from the potential for teacher professional development engendered by the examining process itself. Examiners in other subjects with a practical component attest that involvement in examining work is hugely beneficial to their own teaching practice – it develops their focus on what constitutes high-quality work and brings them into contact with the work of teachers and students in other schools, broadening their perspective on their own practice. This model of assessment would require the involvement of large numbers of examiners, so a substantial proportion of the teaching cohort would experience it.

Likely negative impacts

While there are, no doubt, science classrooms in which not enough practical work is being done, we should not lose sight of the fact that there are many classrooms where good practical work *is* being done. This is because many science teachers recognise that, to a large degree, this is what science is all about, and also recognise that the right blend of practical and theoretical work is the best way to master the discipline, irrespective of how this mastery will subsequently be assessed. It would be a cause of concern if any new practical examination component were to lead to these learning experiences becoming worse rather than better.

Two aspects of the assessment in particular have the potential for negative impact. First, the practical assessment is, of necessity, an individual one. This was mentioned in section 6.9.1 as a potential source of inauthenticity in the tasks, but is mentioned here in the context of its potential impact. It is clear from the responses of teacher participants in the trial that, if this model of assessment were to be rolled out, they would seek to have students engage in individual practical work, as distinct from group practical work, to a greater degree, even if not to the exclusion of group-work. As things stand, collaborative learning is probably less prevalent in schools than most stakeholders would wish for, and science practical work is one of the few areas where it is the norm. So, if reducing the amount of collaborative learning were to be a consequence of any policy intervention, this would be regarded as a negative impact, notwithstanding the value one might place on student independence.

However, perhaps this deficiency should not be overplayed. It is by no means necessary for students to have engaged with a large amount of individual practical work in order to succeed in an assessment of the type proposed. Individual practical work is more difficult for teachers to organise, and practical work in groups will continue to be an extremely effective way for students to learn. The use of an individual assessment at the end of this process simply requires greater focus on ensuring that all members of the group fully participate, that they all share in the opportunities to develop all of the constituent skills required, and all have opportunities to 'take the lead' at significant points in the work. These are characteristics of well-organised group work anyway. Coupled perhaps with occasional opportunities for individual work, it would be a highly effective means of preparing for an individual practical assessment.

A second potential negative consequence arises from the nature of the tasks, as discussed in 9.1 and 9.2 above. If the tasks test the execution of comparatively well-rehearsed skills in relatively familiar contexts, will those teachers who currently use practical work as a vehicle for open and creative investigative work feel compelled to narrow its focus? Does assessing *some* of what practical work is about lead to *the rest* of what practical work is about getting less attention than if we do not assess practical work at all? This is speculation, and one might equally argue that teachers who currently engage in such work will feel more supported and vindicated than heretofore, by virtue of the fact that practical work is being assessed at all, and will be all the more enthused. Nonetheless, the question warrants consideration.

Neutral or uncertain impacts

It is difficult to assess the impact that the proposed model of practical assessment will have on student attitudes to science. The responses from students involved in the trial to questions

that probed this area were positive but perhaps not quite as positive as might have been expected. (See Section 4.1.3 and Appendix D.) The great majority of students enjoyed participating in the trial and agreed quite strongly that they would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate. However, the strength of agreement was less pronounced about whether practical skills *should* be tested as part of the Leaving Certificate examination and less strong again about whether this was a fair way to do it.

We must exercise some caution as to the degree of representativeness of these voices, as the schools participating in the trial all volunteered to do so. It is likely that the sample is biased towards schools where practical work is valued, routinely done, and probably done well, so the voices of students in schools where practical work is not valued, not done, or done poorly, are almost certainly absent, and yet these are the students likely to be most affected by an intervention that increases the prevalence of practical work.

Would the introduction of this form of practical assessment lead to more positive student attitudes to practical work and to science generally? To the extent that it would increase the amount of practical work done and the emphasis placed on developing the associated skills, one would certainly be inclined to think so. Less certain, though, for students already experiencing good practical work is the effect of a possible change of mind-set from '*we are doing this because it is how you learn and do science'* to '*we are doing this because we will be tested on how well we can do it'*. For some, having something tested takes the joy out of it, while for others, it increases their interest, enthusiasm, and application. Also, any shift away from group-work and towards individual work that does emerge is likely to be seen as increasing the attractiveness of the subjects to those who like to work alone and decreasing it to those who like to work with others.

9.4 Practicalities of implementation

Among the questions to be addressed by the trial were: *What delivery models might be feasible, given systemic constraints, including examiner supply?* and *What resources (financial and other) would the SEC require in order to implement a full rollout of such a component under any feasible models identified?* Chapter 8 laid out the very significant challenges that would be faced if this model of assessment were to be rolled out, and some of the associated costs. The SEC currently has considerable difficulty arranging for the conduct of the existing oral and practical examinations, particularly with regard to examiner supply, and this assessment model would be even more challenging than the existing ones in that respect, particularly given the twin needs of the subject teacher needing to be available while at the same time those teachers forming the pool from which examiners are predominantly to be drawn. Additionally, the disruption the assessment would be causing to schools, while not excessive, needs to be acknowledged, along with the level of preparation required of teachers. Accordingly, it would seem unwise to seek to simply drop the proposed model on top of an already overly burdened system, without considering other options.

Notwithstanding this, if a decision is made that this assessment is sufficiently valuable to outweigh the difficulties, the problems should not be regarded as insurmountable, and a number of possible models were outlined in Chapter 8. Also, given the existing difficulties with delivery of oral and practical examinations under the current arrangements, it may be

opportune to stand back and review the entire model of implementation of oral and practical examinations across all subjects. Such a review should include consideration of making changes to the way in which the school year itself is organised. If such a broader review were to happen, then the science practical examinations could certainly be considered as part of this broader review, and it might become a more tractable problem.

As noted in Chapter 8 and elsewhere, it would not be safe to assume that laboratory facilities and equipment in schools are currently of a sufficient standard to support a rollout of this model of practical assessment without further investment, the scale of which remains to be determined.

Two more of our initial questions might be regarded as practicalities of implementation: *Can examiners assess twelve candidates in a single session?* and *Can three 90-minute examination sessions be accommodated during a school day?* While the trial did not throw up any specific concrete difficulties in centres where examiners had 12 or close to 12 candidates, the feedback from examiners suggested that they found it difficult to do what was required of them with this many, and they suggested fewer. Likewise, the degree to which students indicated a belief that examiners were too busy to observe them properly is noteworthy, especially since public confidence in the certificate examinations requires a high level of stakeholder support. It would therefore seem wise, if the assessment is to be rolled out, to restrict the maximum number of candidates per session to approximately eight. This affects the logistics and costings, as detailed in Chapter 8.

While feedback from examiners also suggests that it will be difficult to manage three sessions in one day, we consider this still to be a feasible target to aim for, especially since a reduction to two per day, combined with reducing the number of candidates per session, would lead to a considerable extension of the period of disruption to schools. However, it is clear from the experience of the trial that it is not feasible to have three sessions on the first day in any school, as there are a number of matters that need to be attended to by the examiner, with the assistance of the DST in some cases, in advance of the first session.

This question of the number of sessions sustainable per day brings us to another of our initial questions: *What level of intrusion on the other work of the school would such an examination component have?* The feedback from principals and teachers in Chapter 4 is germane. The level of disruption would be considerable for the period of time in which the assessment is happening – more so, for example, than the disruption caused by the oral examinations, but should probably not be considered unreasonable in the context of the accepted importance of practical laboratory skills in these subjects.

9.5 Stakeholder support

The only one of our initial questions that we have not explicitly revisited in the preceding sections of this chapter is: *To what extent would such an examination component receive the support of students, teachers, and school authorities?* The answer to this is interlinked with all other considerations. It is clear that the most significant stakeholders see great value in assessing practical skills in science and are very positively disposed to it in principle. Nevertheless, when it comes to the details of implementation, many see significant difficulties, and may have conflicting views about how those might be resolved. Certainly, the value that

the science community and the science education community attach to practical work will predispose many stakeholders towards strong support. Nevertheless, maintaining this will require that the practical assessments are seen to be fair, that they are accepted as testing what they are supposed to be testing, that they have a significant positive impact on the learning experiences of students, and that they are arranged to cause as little disruption as possible.

Another factor that has not been mentioned thus far, but which potentially impinges on stakeholder support, is the fact the mark for the practical skills element, having been awarded in real time on the basis of observation of a live performance, cannot feasibly be made amenable to appeal. There are certain elements of other Leaving Certificate examinations that are, for similar reasons, not open to appeal, so this need not rule out this assessment mode. However, the lack of availability of an appeal mechanism for any element of a public service decision-making process has to at least be recognised as a weakness. While the great majority of stakeholders will understand why this element cannot be appealed, it is nonetheless likely to be a source of concern, as there will undoubtedly be some candidates every year who believe they were not awarded the mark they deserved.

9.6 Assessing the value against the costs

All of the potential benefits of introducing the proposed model need to be set against all of the potential costs, (financial and other). The decision to be made is not a straightforward one, but it is hoped that this report has laid out clearly all the issues that require consideration when making the overall evaluative judgment required. In the preceding sections of this chapter, we have looked back to all of the initial questions set out in Section 1.4 at the start of the report, and drawn attention both to the information provided by the trial and other relevant considerations when addressing each.

The trial has demonstrated that this form of practical assessment can successfully be used to measure the skills that it set out to measure, in line with the purpose laid out for it in the subject specifications. However, it is clear that there are some who would consider that this was an overly limited purpose, and, given that this assessment model cannot measure what they consider to be more important skills, an alternative means of assessment should be sought. Notwithstanding this reservation, the absence of any form of direct and meaningful assessment of practical skills as part of the Leaving Certificate examination in these science subjects is a long-running source of considerable dissatisfaction. Practical assessment was proposed at the time of syllabus changes in 1982 but not implemented. Fifteen years later, the then Department of Education and Science funded a feasibility study of another model of practical assessment in Leaving Certificate Chemistry and Physics (Report, 1997), and shortly thereafter the report of the Task Force on the Physical Sciences (2002) recommended that the introduction of a practical assessment component should proceed ‘as quickly as possible’. Following the preparation in 2013 of the three draft subject specifications by the NCCA, there is certainly a strong expectation in the system that a practical component will be introduced. For instance, documents such as the report of the Stem Education Review Group (2016) noted this development. There will be considerable disappointment, to say the least, if some form of assessment of practical work is not introduced.

Certain alternative models of practical assessment through which it might be feasible to test a more expansive range of skills require, of necessity, the involvement of teachers in a school-based model of assessment, and are therefore not available in the context of the Leaving Certificate examination. External coursework assessment models have the potential to assess a different range of skills, many of which are highly valued, but they will lack the element of direct assessment by observation of student practical skills. There is no doubt that the introduction of the assessment model proposed and trialled here will lead to an increased emphasis on practical work in the teaching and learning of the subjects concerned and a general overall increase in the skill levels of students in carrying out such practical work. Alternatives are unlikely to have a similar impact, although they may have other positive effects.

All of the challenges of implementation, the financial cost of actually carrying out the assessment being in many ways the least of these, need to be considered. If a decision is made in principle to proceed, the Commission is strongly of the view that a major review is required of how all oral and practical assessment is organised, how it integrates with the other needs of the education system, and how it is to be supported by the education community at large.

References

Adams, R.M., and Murphy, R.J.L. (1982) The Achieved Weights of Examination Components. *Educational Studies*, 8, 15-22

Baird, J. and Scharaschkin, A. (2002) Is the Whole Worth More than the Sum of the Parts? Studies of Examiners' Grading of Individual Papers and Candidates' Whole A-Level Examination Performances. *Educational Studies*, 28, 2, 143-162

Cresswell, M. (1997) *Examining Judgements: Theory and Practice of Awarding public examination grades*. PhD thesis, University of London Institute of Education: London.

Cresswell, M. (2000) Defining, Setting and Maintaining Standards in Curriculum-Embedded Examinations: Judgemental and Statistical Approaches. In H. Goldstein, and T. Lewis, Editors *Assessment: Problems, developments and statistical issues*. (pp. 57 to 84). Chichester: John Wiley and Sons.

Good, F. J. and Cresswell M. J. (1988) Grade Awarding Judgements in differentiated examinations. *British Educational Research Journal*, 14, 3, 263-281

Holman, J. (2017) *Good Practical Science*. London: Gatsby Charitable Foundation. http://www.gatsby.org.uk/uploads/education/reports/pdf/good-practical-science-report.pdf

Millar, D. Kellaghan, T. and Mac Aogáin, E. (2006) *A Study of the Intended and Achieved Weights of Components in the Leaving Certificate Examination*. Dublin: Educational Research Centre

NCCA (2013) *Assessment of Leaving Certificate Science*.

NCCA (2014) *Leaving Certificate Science Subjects – Working Towards Implementation*.

NCCA (2014) *Senior Cycle Sciences – Assessment Outline*.

Newton, P.E. and Shaw, S.D. (2014) *Validity in Educational and Psychological Assessment*. London: Sage publications.

Report (1997) *Feasibility Study on Practical Assessment for Leaving Certificate Physics and Chemistry*. Report prepared by the Steering Committee for the Department of Education and the National Council for Curriculum and Assessment, Dublin.

Scharaschkin, A and Baird, J (2000) The effects of consistency of performance on A Level examiners' judgements of standards, *British Educational Research Journal*, 26, 3, 343-357

STEM Education Review Group (2016) *STEM Education in the Irish School System – A Report on Science, Technology, Engineering and Mathematics (STEM) Education, Analysis and Recommendations*. Dublin: Department of Education and Skills.

Task Force (2002). *Report of the Task Force on the Physical Sciences*. Dublin: Department of Education and Science.

Appendices

Many of the appendices are substantial documents in their own right, so the majority are being made available on request as a separate circulation. Only the appendices that are shorter and/or that readers are most likely to wish to refer to are included here.

Appendix A: Membership of Steering Group and Implementation Group (included)

Appendix B: External reviewer's Terms of Reference, Interim Report, and Final Report (included)

Appendix C: Trial documentation: letters, instructions and forms (separate circulation)

Appendix D: Summary of responses to Student questionnaire (included)

Appendix E: Task booklets, Biology (separate circulation)

Appendix F: Task booklets, Chemistry (separate circulation)

Appendix G: Task booklets, Physics (separate circulation)

Appendix H: Chief examiner report for Biology (separate circulation)

Appendix I: Chief examiner report for Chemistry (separate circulation)

Appendix J: Chief examiner report for Physics (separate circulation)

Appendix K: Acronyms used in this report (included)

Appendix A: Membership of Steering Group and implementation Group

Steering Group

The membership of the steering group was adjusted over the course of the project, so not all members listed below were necessarily members for the full duration.

- Hugh McManus, Assistant Head of Examinations and Assessment, State Examinations Commission (Chair)
- Andrea Feeney, Director of Operations, State Examinations Commission
- John Hammond, Chief Executive, National Council for Curriculum and Assessment
- Barry Slattery, Director of Curriculum and Assessment, National Council for Curriculum and Assessment
- Eamonn Moran, Principal Officer, Curriculum and Assessment Policy Unit, Department of Education and Skills
- Ciara Molloy, Assistant Principal, Curriculum and Assessment Policy Unit, Department of Education and Skills
- Ruth Richards, Postprimary Inspector, Department of Education and Skills
- Jason Kelly, Postprimary Inspector, Department of Education and Skills
- Brian Power, Principal Officer, Planning and Building Unit, Department of Education and Skills

Implementation Group

- Richard Coughlan, Trial manager (chair)
- Hugh McManus, Assistant Head of Examinations and Assessment, State Examinations Commission
- Fiona Desmond, Examinations and Assessment Manager, State Examinations Commission
- Ruth Richards, Postprimary Inspector, Department of Education and Skills
- Anna Walsh, Education Officer, National Council for Curriculum and Assessment
- Barry Slattery, Director of Curriculum and Assessment, National Council for Curriculum and Assessment

Appendix B: External reviewer's Terms of Reference, Interim Report, and Final Report

Terms of Reference

External Project Review of Trialling of Practical Assessment in Senior Sciences

Context

Given the systemic importance and far-reaching consequences of any decision to roll out a practical assessment as a component of the Leaving Certificate examination in the science subjects, It has been proposed that the trialling project would benefit from being reviewed by an independent external agent, in order to provide assurance to stakeholders that the trialling project is fairly and appropriately planned and executed.

Brief for external reviewer

Given that the trial is at an advanced stage of preparation, the first phase should be carried out as soon as is practicable.

Phase 1

The reviewer will examine relevant documentation related to the trial to assess the following:

- The extent to which the plan is clear as to the project's scope: that it sets out clearly what questions are intended to be answered by the project, and what questions fall outside its scope.
- The extent to which the proposed methodology represents a good way to get answers to those questions.
- The extent to which the detailed arrangements for the plan are fit for purpose.

The reviewer will have access to the trial manager and the AHEAD overseeing the project in order to get any further documentation or information that he or she considers necessary to carry out this assessment.

The reviewer may make any suggestions for amendments to the plan, methodology, or arrangements that he/she considers would improve it and would be viable. It will remain the responsibility of the implementation group and SEC management to decide the extent to which such suggestions can be acted upon.

Phase 2

As more detailed documentation becomes available (such as instruments for capturing and analysing the views of students, teachers, and school authorities) the reviewer will check these and make any suggestions he or she considers appropriate.

The reviewer will observe the implementation of the assessment in one or more schools, if he or she is available to do so.

Phase 3

After implementation, when the draft project report is available, the reviewer will review the draft project report and any other documentation considered appropriate and assess the following:

- The extent to which the plan has been carried out in a rigorous manner in accordance with its principles and specification
- The extent to which any proposed conclusions are valid and based on sound evidence.

The reviewer will again have access to the trial manager and the AHEAD overseeing the project in order to get any further documentation or information that he or she considers necessary to carry out this assessment.

The reviewer will compile a brief report after Phase 1 and Phase 3 of the review, and at any intermediate stage that he or she considers appropriate. These brief reports from the reviewer will be presented in draft form to the AHEAD overseeing the project, who will forward them to the implementation group. The implementation group may seek observations from others involved in the project on aspects of the reviewer's draft reports. The AHEAD will collate the observations of the implementation group and any others consulted and present these to the reviewer. The reviewer will consider these observations before finalising any such report. The reviewer's finalised report following Phase 1 and finalised report following Phase 3 will be presented to the AHEAD overseeing the project, who will ensure that each is presented to the implementation group, the project steering group, SEC management board, and the Commission (board of commissioners).

Interim report

External Project Review of Trialling of Practical Assessment in Senior Sciences

Interim report, 1 December 2017

John Holman

1 My brief

The State Examinations Commission has asked me to provide an independent review of the trial of the new arrangements for assessment of practical skills in the Leaving Certificate for senior science. The purpose of this review is to provide assurance to stakeholders that the trialling project is fairly and appropriately planned and executed.

2 My actions so far

I have received and reviewed documentation:

- The proposed new assessment arrangements
- The application process for trial schools
- The briefing sent to trial schools
- Details of the initial try-out of the practical activities
- The practical tasks set to students for the trial, in the form of the 'live' examination papers
- The Leaving Certificate syllabuses in physics, chemistry and biology.

I have received and provided comments on:

- Follow-up questionnaires for students, school principals, teachers and examiners.

I have had briefings on the trial from

- Hugh McManus, Assistant Head of Examinations and Assessment, SEC
- Richard Coughlan, Trial Manager
- Fiona Desmond, Chief Examiner for Leaving Certificate Chemistry.

I have also had informal discussions with Anna Walshe of the National Council for Curriculum and Assessment, who played a major role in the design of the new assessment scheme.

I have visited schools taking part in the trial: Malahide Community School, Dublin and Fingal Community School, Swords. At Malahide I observed chemistry assessments and at Fingal I observed Physics. At both schools I spoke to the examiner and the designated support teacher.

3 Interim observations about the trial

I was asked to report on the plan for the trial, in particular (quoting from my brief from the SEC):

- *The extent to which the plan is clear as to the project's scope: that it sets out clearly what questions are intended to be answered by the project, and what questions fall outside its scope.*
- *The extent to which the proposed methodology represents a good way to get answers to those questions.*
- *The extent to which the detailed arrangements for the plan are fit for purpose.*

The timing of my review did not allow me to suggest amendments to the methodology. However, I am content that the methodology is appropriate, the project plan is clear and that the detailed arrangements are fit for purpose. However, I do have some comments about the wider context within which the trial is set (see section 5).

On the basis of what I have seen so far, I am confident that the preparation for the trials has been thorough and methodical. From what I have observed at two schools, execution of the trials has been smooth and efficient.

The documentation accompanying the trials is clear and thorough. Schools have been well briefed so they knew what to expect, without revealing to them the tasks that students would be set. The examiners to whom I spoke and whom I observed were confident and well prepared. The experimental tasks were pre-tested to make sure they all work in the conditions that would be encountered in action.

4 Future reporting

My next report will be written when the draft project report is available. My brief requires me to:

review the draft project report and any other documentation considered appropriate and assess the following:

- *The extent to which the plan has been carried out in a rigorous manner in accordance with its principles and specification*
- *The extent to which any proposed conclusions are valid and based on sound evidence.*

I am assuming that the draft project report will include tables showing data from the marked scripts, collations of the questionnaire results, etc. In addition, in order to assess whether such summary information fairly reflects the underlying evidence, I will need access so some of the primary sources of evidence, including:

1. Samples of the questionnaires (students, school principals, teachers and examiners), selected to provide a range of opinions
2. Sample scripts across biology, chemistry and physics, selected to provide a range of marks, and the corresponding examiners' scoring sheets
3. Any other primary evidence that SEC considers appropriate.

Ideally, I would have access to all the original documents so I can do my own sampling.

5 Some comments on the wider context

My brief requires me to report on the conduct of the trial itself, and to provide quality assurance on the trial. This I will do.

However, I would like to offer some reflections on the wider context within which this trial is taking place.

5.1 Why assess practical skills and knowledge in science?

Experimentation gives science its identity. Science uses experiments to discover the realities underlying the world, and this practical approach is as intrinsic to young learners as it is to professional researchers.

The international study that I carried out for the Gatsby Foundation, *Good Practical Science*[15], finds five purposes for carrying out practical work in science education.

A. to teach the principles of scientific enquiry;

B. to improve understanding of theory through practical experience;

C. to teach specific practical skills, such as measurement and observation, that may be useful in future study or employment;

D. to develop higher level skills and attributes such as communication, teamwork and perseverance;

E. to motivate and engage students.

Not all of the outcomes from these five purposes can or should be assessed. But there is a strong case for assessing understanding of the principles of scientific enquiry, and for assessing practical skills. The first of these can to some extent be assessed through written questions, but the only valid way that practical skills can be assessed is through direct and indirect assessment of individual students' own practical work. So there is a strong case that, if science is accepted to be a practical subject, practical assessment should be an intrinsic part of the assessment model.

There is another reason for practical assessment. In a high-stakes assessment such as the Leaving Certificate, it has to be accepted that to a large extent, *what gets taught is what is assessed*. From conversations with people associated with this trial, I judge that a desire to get schools doing more practical science is an important part of the motivation for it.

5.2 Relation of practical assessment to the Leaving Certificate syllabuses

The practical assessment in any qualification needs to relate to (1) the subject matter (2) the assessment objectives, in the syllabus. Having reviewed the Leaving Certificate science syllabuses, I note that the Objectives for Skills include skills such as 'perform experiments safely and collaboratively'; 'select and manipulate suitable apparatus ….';

[15] Good Practical Science, the Gatsby Foundation, 2017.
http://www.gatsby.org.uk/education/programmes/support-for-practical-science-in-schools

'make accurate observations and measurements'[16], that can only be validly assessed in the context of students' own practical work. They cannot be validly assessed by written questions alone. This underlines the importance of introducing practical assessment into the science Leaving Certificate, which is the motivation behind this trial.

Having reviewed the assessment tasks and questions used in the trial, I note that they are outdated in some of their subject matter – for example, the chemistry tasks include the action of water on calcium carbide and the 'brown ring' test for nitrates – both of which have disappeared from most modern chemistry syllabuses because they are no longer relevant to modern chemistry.

What is more, in all three subjects, the tasks themselves mainly require students to carry out a set of highly specific instructions, while being observed by the examiner. My judgement is that the skills being assessed relate to manipulation and observation within a limited range of practical situations, and to the interpretation of the results in a set of written answers. The tasks in the trial do not require students to design and plan an experiment in a new situation in a way that would assess their understanding of the principles of scientific investigation, which is one of the aims of the Leaving Certificates in science.

Having reviewed the current Leaving Certificate syllabuses, I understand the limitations acting on the designers of the tasks. The tasks have to reflect the expectations of the syllabus on which students have been taught. These expectations are mainly defined by the subject content and by the mandatory experiments[17], together with the stated Objectives.

However, I understand that new syllabuses for the science Leaving Certificates are to be introduced shortly. There will be opportunities, with these new syllabuses, to include modern subject content and design practical tasks that are based on this content and which – if the Objectives call for it – assess a wider range of practical skills. This is not a criticism of the current trial *per se*: indeed it makes good sense to test the principle of a practical assessment based on the current syllabus in order to build confidence to include practical assessment when the new syllabuses are introduced.

5.3 Models of assessment of practical science

In my international study I found a variety of different models for assessing practical science. These include:

A. Direct assessment by teachers, in which the class teacher observes students' practical work and assigns a grade, often based on defined criteria. This approach may be accompanied by moderation to align the assessment by different teachers in the same school, and between different schools. This is the commonest approach across the countries I have visited and surveyed.

[16] These are taken from the chemistry syllabus, but there are similar statements in biology and physics.
[17] In the physics and chemistry syllabuses, mandatory experiments are clearly specified. However, it is less clear to me in the biology syllabus which experiments are mandatory.

B. Direct assessment in a practical examination, in which an examiner observes and assesses students' work, combined with indirect assessment of their results. This is the model on trial in Ireland.
C. Endorsement by the school that students have carried out a specified range of practicals, without assigning a grade. This is the model that is being moved towards in England.

In my view, the most valid model for practical assessment is (A), direct assessment by teachers. In a system where teacher assessment is the trusted norm, this method can be secure and reliable. However, I recognise that, for reasons specific to the Irish teaching context, teacher assessment is not a realistic prospect.

Final report

External Project Review of Trialling of Practical Assessment in Senior Sciences

Final report, 21 May 2018

John Holman

1 My brief

The State Examinations Commission has asked me to provide an independent review of the trial of the new arrangements for assessment of practical skills in the Leaving Certificate for senior science. The purpose of this review is to provide assurance to stakeholders that the trialling project is fairly and appropriately planned and executed.

In my interim report of December 2017, I reported on the plan for the trial, in particular (quoting from my brief from the SEC):

- *The extent to which the plan is clear as to the project's scope: that it sets out clearly what questions are intended to be answered by the project, and what questions fall outside its scope.*
- *The extent to which the proposed methodology represents a good way to get answers to those questions.*
- *The extent to which the detailed arrangements for the plan are fit for purpose.*

The timing of my review did not allow me to suggest amendments to the methodology. However, I reported that I was content that the methodology was appropriate, the project plan clear and the detailed arrangements fit for purpose.

In preparation for this, my final report, I have read in detail the SEC's draft *Report on the Trialling of the Assessment of Practical Work in Leaving Certificate Biology, Chemistry and Physics*, and I have revisited the evidence collected for my Interim Report. I now report on:

- *The extent to which the plan has been carried out in a rigorous manner in accordance with its principles and specification.*
- *The extent to which any proposed conclusions are valid and based on sound evidence.*

2 Sources of evidence

I received and reviewed the following documentation:

- The proposed new assessment arrangements;
- The application process for trial schools;
- The briefing sent to trial schools;
- Details of the initial try-out of the practical activities;
- The practical tasks set to students for the trial, in the form of the 'live' examination papers;
- Sample examination scripts for candidates in physics, chemistry and biology. I reviewed 23 scripts, some of which had been double-marked;
- The Leaving Certificate syllabuses in physics, chemistry and biology;

- The SEC's draft *Report on the Trialling of the Assessment of Practical Work in Leaving Certificate Biology, Chemistry and Physics* (dated 27 April 2018).

I received and provided comments on:

- Drafts of the follow-up questionnaires for students, school principals, teachers and examiners.

I had briefings on the trial from

- Hugh McManus, Assistant Head of Examinations and Assessment, SEC
- Richard Coughlan, Trial Manager
- Fiona Desmond, Chief Examiner for Leaving Certificate Chemistry.

I have also had informal discussions with Anna Walshe of the National Council for Curriculum and Assessment, who played a major role in the design of the new assessment scheme.

I visited schools taking part in the trial: Malahide Community School, Dublin and Fingal Community School, Swords. At Malahide I observed chemistry assessments and at Fingal I observed Physics. At both schools I spoke to the examiner and the designated support teacher.

At all times Hugh McManus has willingly and openly responded to my queries.

I have found the documentation clear and easy to follow. In particular, the SEC's draft Report is exemplary in its clarity. Assessment is a complex technical field which the authors have rendered concise and accessible.

3 Review of the draft Report

In this section I briefly review the draft report, chapter by chapter.

Chapter 1 Context, purpose and scope of the trial These are set out concisely.

Chapter 2 Planning and implementing the trial. I commented in my Interim Report that the trial methodology was appropriate, the details were clearly communicated and the arrangements were fit for purpose. Chapter 2 confirms my earlier view and shows that the plan has been carried out in a rigorous manner in accordance with its principles and specification. This is in spite of the arrival of severe weather in the form of Hurricane Ophelia, which could have blown the trial off course, but did not.

In any trial of this kind, it is essential that the intentions and arrangements are communicated clearly to schools, and this was done with consistent clarity.

I note that the sample of schools was partially self-selected, which (as the report acknowledges) makes it likely that the sample was skewed towards schools whose facilities and teachers dispose them more favourably towards practical work.

I note that the original intention was for the new assessment arrangements to begin in September 2018, but that this is no longer feasible. I understand that the new syllabuses for the science Leaving Certificates are now to be introduced from 2021, and it would seem sensible that, if a decision is taken to proceed with the practical

assessments, this should be from 2021 so the assessments are consistent with the new syllabuses.

As I noted in my Interim Report, some of the assessment tasks are outdated in their subject matter – for example, the chemistry tasks include the action of water on calcium carbide and the 'brown ring' test for nitrates – both of which have disappeared from most modern chemistry syllabuses because they are no longer relevant to modern chemistry. I trust that when the specifications are updated in 2021, outdated practical science of this kind will be replaced.

Chapter 3 The practical assessment in detail. This chapter describes how the assessments were carried out in practice, and I can confirm that the account corresponds to what I saw in my own (admittedly limited) observations in two schools. The chapter shows that the original intentions were carried out in practice.

Chapter 4 Feedback from participants. This chapter is based on an analysis of questionnaires administered after the trial. I was asked to comment on the draft questionnaires, which I believe were amended following some of my comments. Although I did not inspect any of the completed questionnaires, I believe their analysis was carried out professionally and robustly. The inclusion of quotations from participants, along with a full quantitative analysis, gives a textured picture of participants' responses to the trial.

The analysis in this chapter provides valuable evidence which has been used in Chapters 8 and 9 to evaluate the trial and its future consequences.

Chapter 5 Evaluating the technical quality of the assessment – quantitative evidence. This impressive chapter provides a clear and expert analysis of the assessment results, including robust statistical analysis. It lays solid foundations for the conclusions in chapters 8 and 9.

Chapter 6 Evaluating the technical quality of the assessment – qualitative evidence. This chapter gives subject-specific and task-specific detail of the assessments in practice. It will be particularly valuable to those who have the responsibility for implementing the arrangements in practice, if the decision is taken to roll out.

Chapter 7 The digital trial. This chapter summarises the results of the small scale trial to investigate the feasibility of digital assessment of practical work. For various reasons, including technical difficulties and the small scale of the trial, I do not believe any conclusions can be drawn from it.

Chapter 8 Rolling out the model nationally – scalability, cost, and scheduling. The chapter provides a robust and honest analysis of the implications of roll-out, from a practical and a financial point of view, and it is a strong basis for the policy conclusions in chapter 9. An important conclusion is that assessment of groups of 12 is very challenging, which I noted in my own observations in pilot schools. I endorse the suggestion that eight should be the upper limit of group size, noting that the effect of such a change on the overall cost would be smaller than might be expected.

Chapter 9 Towards a policy decision on roll-out. I comment on this chapter, in greater detail than the others, in the next section.

4 Towards a policy decision on roll-out

Chapter 9 provides an excellent discussion of the benefits and costs of rolling out the practical assessment in all schools in Ireland. It includes an expert discussion of some of the technical aspects of the assessment model in terms of validity and fairness, and it considers alternatives. Its conclusions are valid and based on sound evidence.

The chapter is even-handed and discusses positive and negative aspects of the trial without bias.

Inevitably any discussions of positive and negative effects will find it difficult to give weighting to the various factors. In the following, I highlight the factors which seem to me the most significant in making the decision on roll out. In doing so, I may be going beyond my brief, but I hope my comments may be helpful.

The discussion of validity in 9.1 is excellent and based firmly on the evidence of the trial. I agree with the suggestion that the best way of dealing with the issues of validity and security is to **use unfamiliar tasks** (as opposed to tasks selected from the mandatory experiments). The report considers the advantages and disadvantages of different ways of doing this and I agree that the only fair way is to give all candidates equal knowledge of the tasks that will be used. While this is achievable, it should not be underestimated how much thought and work will be needed to construct a suitable bank of items.

The analysis in 9.2 concludes that

*'… it will **not** be possible with this model to test the candidates' capacity to apply their knowledge or skills to unfamiliar contexts. Those cognitive skills will have to remain in the target domain of the written examination only. This is consistent with the purpose of the assessment as expressed in the subject specifications themselves, but not with the purposes expressed in some of the ancillary documentation.'*

I take this to mean that the purpose of the practical assessment remains the assessment of practical skills rather than cognitive ones, and this seems realistic and acceptable.

The discussion in 9.3 of **Impact on teaching and learning** is important. While it could be argued that the purpose of introducing practical assessment is to do just that – assess practical skills – experience in England and elsewhere makes clear that changes to assessment have powerful effects on the way a subject is taught. It is my understanding that an important part of the motivation for introducing practical assessment is to drive an increase in the quantity and quality of practical science in schools in Ireland. It should not be underestimated how powerful this effect can be. However, I understand that there is wide variation in the quality of practical science facilities in Irish schools, and this will of course need to be taken into account. The most desirable outcome would be that the introduction of this assessment will drive a

levelling of laboratory facilities in schools, though this has clear implications for resources.

The outcome to be avoided is one where the practical assessments tasks set each year are so predictable that teachers are tempted to concentrate their practical science lessons narrowly on preparation for a limited number of predictable assessment tasks. This would be a waste of the opportunity to drive wide improvement of practical science, and it is one reason why I favour an approach that uses unfamiliar tasks as far as possible.

5 Some further observations

Disruption

The amount of disruption caused by the preparation for, and execution of, these assessments should not be underestimated, particularly in a country where few schools have laboratory technicians to help. The report pays careful attention to such disruption, and I agree with the conclusion that it is acceptable in the light of the importance of practical skills in science.

External examining as professional development

The model of assessment proposed uses teachers from other schools to act as external examiners, and there is discussion of whether sufficient numbers will be available. I endorse the finding that participation in external examining is an excellent form of professional development for teachers. It provides an opportunity for teachers to see how things are done in other schools and to have professional interaction with other teachers, both formally through the assessment procedures and informally through interactions around the fringes of the assessment process itself. These interactions help to build a community of professional practice and their benefits should not be underestimated.

Such considerations should be factored in when considering the costs and benefits of a system which takes external examiners away from their own schools during the assessment season.

Should the same approach be taken for all sciences?

The assumption behind the trial is that the same model can be validly used to assess all three of biology, chemistry and physics. A common set of assessment objectives is used, and these can be flexed to be appropriate for all three sciences, although in the case of biology they will inevitably skew the tasks towards those that are apparatus-based rather than field-based.

Yet the nature of the three sciences is quite different, particularly in the way they approach the design and planning of investigations. Section 6.10 discusses the content validity of the range of biology tasks, and suggests that a coursework model of assessment may be more appropriate. While I have some sympathy with this view, it seems to me that the scale of the change being contemplated across all sciences in all schools, means that for practical purposes it would be best to use a single model of assessment across all three sciences.

Is a wide range of practical skills being assessed?

In my Interim Report, I noted: '….. *the tasks themselves mainly require students to carry out a set of highly specific instructions, while being observed by the examiner. My judgement is that the skills being assessed relate to manipulation and observation within a limited range of practical situations, and to the interpretation of the results in a set of written answers. The tasks in the trial do not require students to design and plan an experiment in a new situation in a way that would assess their understanding of the principles of scientific investigation, which is one of the aims of the Leaving Certificates in science.'*

It remains my view that this model of practical assessment cannot assess in a valid way students' skills in designing a scientific investigation. There is simply not enough time within this model to design, plan and carry out an investigation, which typically requires several lessons to do in a valid way. So if *understanding of the principles of scientific investigation* remains one of the aims of the Leaving Certificate (and I hope it will) it will have to be assessed through the written papers rather than the practical assessment.

The weighting of practical assessment

It is proposed that the practical assessment should contribute 30% of the whole. This proposal seems to have the approval of trial participants, but experience suggests it may be a little on the high side. The higher the weighting given to practical, the greater the impact, both positive and negative, on teaching in schools. If one of the concerns about the practical assessment proposals is that they will lead teachers to concentrate excessively on a narrow range of practical activities, one way of mitigating this risk would be to reduce the weighting from 30% to, say, 25%.

Experience in England

England has a long experience of assessment practical science at Leaving Certificate level, stretching back to the introduction of A levels in the 1950s. Three general models have been tried.

Practical examinations, where tasks are set externally, carried out under examination conditions and marked externally.

Coursework, whereby practical work is assessed by the classroom teacher, usually with some form of external moderation.

Endorsement, whereby the school validates that each student has carried out a specified range of experiments, but no grade is assigned to the students. This approach, introduced from 2015, is currently in use.

Each of these models has its advantages and disadvantages, which have been keenly debated, though it is too early to properly evaluate the endorsement model. My conclusion from these experiences is that it is difficult to identify an objectively 'ideal' model, because each model is so dependent on the policy and cultural context within which it operates. For example, the move away from a coursework model in England

was driven by the realisation that, in an intensely high-stakes assessment environment, teachers were under enormous pressure to award high marks.

I conclude that Ireland's decision as to whether to roll out the trial should be taken in the light not only of the technical validity and reliability of the assessment instruments but also of the practical policy environment.

6 Summary and conclusion

From my own direct observations, from reading background documentation and the SEC's draft report, I conclude that the trialling project was fairly and appropriately planned and executed. Indeed, the quality of planning, execution and reporting were at the top end of my range of expectation. This has been an exemplary exercise from which other policymakers can learn.

Although they are outside my brief, I make the following key observations.

1 The impact of introducing practical assessment on practice in schools should not be underestimated. It will stimulate more practical work, but the maxim 'what is taught is what is assessed' may apply.

2 The DES will need to bear in mind the range of practical facilities available across Ireland, and that schools that took part in the trial are likely to be those with better facilities.

3 Assessing groups of 12 students per examiner is too demanding and should be reduced.

4 Any new practical assessment arrangements need to be fully integrated into the new specifications to be introduced in 2021, in terms of both subject content and the range of expected practical skills.

5 A 30% weighting for the practical assessment seems a little high.

John Holman, University of York, 21 May 2018

Appendix D: Summary analysis of student data from questionnaires

Biology: 50 student questionnaires		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1.	I enjoyed participating in the trial.	6	26	10	4	2
2.	This was a good way to test my practical science skills.	13	27	2	5	1
3.	The test was a fair way of assessing different student science skills	6	17	16	7	1
4.	Practical science skills should be tested as part of the Leaving Certificate.	16	18	3	5	6
5.	I would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate.	23	16	4	3	2
6.	I knew what to expect on the day of the Biology trial.	9	18	11	10	2
7.	I had covered all the background theory work needed for the Biology trial	7	24	6	10	3
8.	The examiner's talk before the Biology practical session(s) was helpful.	9	29	11	1	0
9.	I understood what my Biology task asked me to do.	11	27	4	6	2
10.	My Biology task was too difficult.	5	5	12	19	8
11.	I had enough time to complete my Biology task.	13	26	2	5	1
12.	I had time to spare.	9	25	2	10	4
13.	The examiner was too busy to observe me properly.	3	8	9	23	6
14.	Everything I needed to do my Biology task was available in the lab.	15	27	1	5	2
15.	I had no trouble finding the things I needed for my task.	8	19	10	9	4
16.	Five minutes was enough clear-up time.	10	24	3	8	4
17.	The part of my Biology task I found easiest was: Setting up and gathering the equipment Concluding the experiment and cleaning up					
18.	The thing that I found hardest about my Biology task was: Finding equipment The experiment					
19.	Any other comment about the Biology trial? The perceived inequality of the range of experiments in terms of length and difficulty It was well organised					
20.	Any general comments? The perceived inequality of the range of experiments in terms of length and difficulty Good opportunity for practical learners Having an external examiner was good practice					

Chemistry 50 student questionnaires		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1.	I enjoyed participating in the trial.	6	24	8	10	2
2.	This was a good way to test my practical science skills.	19	22	3	5	0
3.	The test was a fair way of assessing different student science skills	6	16	8	17	5
4.	Practical science skills should be tested as part of the Leaving Certificate.	16	9	12	7	6
5.	I would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate.	19	19	4	4	3
6.	I knew what to expect on the day of the Chemistry trial.	1	25	8	15	0
7.	I had covered all the background theory work needed for the Chemistry trial	7	20	8	14	2
8.	The examiner's talk before the Chemistry practical session(s) was helpful.	4	29	11	5	1
9.	I understood what my Chemistry task asked me to do.	10	26	6	3	4
10.	My Chemistry task was too difficult.	4	9	10	22	5
11.	I had enough time to complete my Chemistry task.	11	17	5	14	3
12.	I had time to spare.	12	16	4	8	8
13.	The examiner was too busy to observe me properly.	10	15	7	15	3
14.	Everything I needed to do my Chemistry task was available in the lab.	13	29	2	2	2
15.	I had no trouble finding the things I needed for my task.	8	17	6	13	6
16.	Five minutes was enough clear-up time.	6	20	1	15	8
17.	The part of my Chemistry task I found easiest was: Setting up: locating and gathering the apparatus Carrying out the tasks Clearing up					
18.	The thing that I found hardest about my Chemistry task was: Finding the correct chemicals Cleaning up afterwards					
19.	Any other comment about the Chemistry trial? The perceived inequality of the range of experiments in terms of length and difficulty It was well organised					
20.	General comments Either the amount of students present for assessment needs to be decreased or more examiners are required to be present. The examiner was too busy. It was well organised and helpful but it needs more organisation					

	Physics 51 Student questionnaires	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1.	I enjoyed participating in the trial.	7	30	5	7	2
2.	This was a good way to test my practical science skills.	14	29	3	3	2
3.	The test was a fair way of assessing different student science skills	3	15	10	17	5
4.	Practical science skills should be tested as part of the Leaving Certificate.	10	16	14	4	7
5.	I would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate.	28	14	3	2	4
6.	I knew what to expect on the day of the Physics trial.	6	27	9	7	2
7.	I had covered all the background theory work needed for the Physics trial	7	28	6	7	3
8.	The examiner's talk before the Physics practical session(s) was helpful.	5	28	15	2	0
9.	I understood what my Physics task asked me to do.	14	27	6	3	0
10.	My Physics task was too difficult.	1	4	9	25	12
11.	I had enough time to complete my Physics task.	22	24	2	1	1
12.	I had time to spare.	14	22	6	6	2
13.	The examiner was too busy to observe me properly.	6	10	14	18	2
14.	Everything I needed to do my Physics task was available in the lab.	18	30	1	2	0
15.	I had no trouble finding the things I needed for my task.	12	25	8	4	2
16.	Five minutes was enough clear-up time.	14	29	2	4	1
17.	The part of my Physics task I found easiest was: Locating and gathering the apparatus or equipment Setting up the apparatus Writing up the experiment Cleaning up afterwards					
18.	The thing that I found hardest about my Physics task was: Setting up the equipment correctly especially momentum experiment Calculating results Graphing the results Measuring Finding the correct equipment					
19.	Any other comment about the Physics trial? The perceived inequity of the range of experiments in terms of length and difficulty Some experiments are difficult for one person to do on their own It was carried out in a very professional manner. Working alone may be difficult in regards to some experiments. The difficult nature of waiting for the examiner to observe each steps or having to repeat a step because the examiner did not observe it					
20.	General comments The perceived inequity of the range of experiments in terms of length and difficulty					

Student questionnaires – composite of replies to questions 1 to 5 across Biology, Chemistry, and Physics questionnaires	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I enjoyed participating in the trial.	7	30	5	7	2
This was a good way to test my practical science skills.	14	29	3	3	2
The test was a fair way of assessing different student science skills	3	15	10	17	5
Practical science skills should be tested as part of the Leaving Certificate.	10	16	14	4	7
I would pay more attention to practical work in class if it were to be tested as part of the Leaving Certificate.	28	14	3	2	4

Appendix K: Acronyms used in this report

AHEAD	Assistant Head of Examinations and Assessment Division
DES	Department of Education and Skills
DST	Designated Support Teacher
EAD	Examination and Assessment Division
EAM	Examination and Assessment Manager
NCCA	National Council for Curriculum and Assessment
SEC	State Examinations Commission