The impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences

Sarah Old

A Research & Development Project

Submitted for the MSc Educational Assessment 2022

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via Oxford Research Archives where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be inserted in the library copy of the dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Old
First Name	Sarah
Faculty Board	Education
Title of Dissertation	The impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via Oxford Research Archives where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

The impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences

Contents

Abstract 3

Introduction 5

Practical work in science and its purposes 5

The assessment of practical work in science 6

The assessment of practical work in the A level sciences in England 8

The impact of the COVID-19 pandemic on practical work in schools and colleges 13

Conceptions of fairness 14

Fairness of A level science practical science assessment during the pandemic 19

Research questions and hypotheses 23

Method 27

Survey 27

Interviews 29

Ethical issues 30

Survey and interview questions 31

Analysis of survey and interview data 33

Results 38

The number of hours of practical work per week 38

The number of hours of practical work and school type, state or independent 41

The focus on the requirement for 12 practical activities 44

The type of practical work completed by students 47

The fairness of the assessment of practical skills in the A level sciences 50

Fairness and the hours of practical work per week 55

Measuring Nisbet and Shaw's typology of fairness 58

Discussion 61

Impact of the pandemic on the delivery of practical work 61

Impact of the pandemic on the fairness of the assessment of practical skills 67

Fairness in the circumstances of the pandemic 72

Fairness, validity and reliability 74

The application of fairness to policy and practice 77

Conclusion 79

Limitations 79

Further research 80

References 82

Appendices 87

Appendix A - Results from additional analysis 87

Appendix B - Results from additional analysis 89

Appendix C - Results from tests of assumptions 93

Appendix D - Results from analysis to show extreme kurtosis on fairness scales 94

Appendix E – Exploratory factor analysis on fairness items 95

Appendix F – CUREC approval 97

Appendix G – Survey questions 98

Appendix H – Interview questions 104

Abstract

The assessment of A level science practical skills is an important part of the qualifications and has been approached in a number of different ways since the qualifications began in 1951. This has included coursework, practical examinations and the current approach of an endorsement to the overall A level grade indicating competence in practical skills. The global COVID-19 pandemic that began in 2020 caused widespread disruption to education including partial school closures and students had their learning transferred to remote means. There was widespread discussion about the impact of the pandemic on the fairness of high-stakes assessments in 2020 and 2021 when public exams were cancelled. This discussion continued for assessments in 2021-22 even though exams did take place in summer 2022.

This study aimed to investigate the impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences. This investigation included research into the impact of the pandemic on the delivery of A level practical work and the fairness of any impacts.

Analysis using appropriate statistical methods showed that partial school closures and remote learning prevented teachers from delivering the same amount and type of practical work as normal. In addition, public health restrictions further affected the ability of teachers to deliver practical work. The method of assessment in light of the cancellation of exams in 2021, Teacher Assessed Grades (TAGs), resulted in an inevitable focus on theory work that could be assessed by teachers to provide the required evidence for TAGs. The rules for the assessment of A level science practical skills were changed to

facilitate the award of the Practical Endorsement in situations where it was not possible for students to complete all the required practical activities as long as they had met the Common Practical Assessment Criteria (CPAC). Analysis showed that survey respondents considered this approach to be fair in a broad sense even if students did not have the same opportunities for practical work as other students in different schools and colleges or in their year of assessment compared with those in other years. This was because the circumstances of the pandemic were no-one's fault and no-one set out for it to be unfair. Teachers had done their best in these circumstances. Further consideration of fairness in light of Nisbet and Shaw's typology of fairness (2020) allowed the interpretation that it was not possible to fully distinguish between the six senses of fairness described and that survey respondents and that the senses of fairness, where identified, were not differentially impacted over the course of the pandemic.

It is concluded that the COVID-19 pandemic did impact on the delivery of A level science practical work and its assessment. The amount and type of practical work that provided evidence for the assessment of the CPAC varied between schools and colleges and also between years of assessment. Any impact on fairness is considered to be acceptable as the changes to the rules for the assessment allowed these differences. The circumstances of the pandemic could not be helped and teachers did their best in often challenging circumstances. This view did not allow for the full identification of the senses of fairness in Nisbet and Shaw's typology (2020) and there was not differential impacts where differences were identified.

Introduction

Practical work in science and its purposes

It is accepted across the world that student participation in practical work is integral to teaching and learning in science in the same way that speaking is key to the learning of modern foreign languages (Gatsby, 2017). The term 'practical work' is used by Abrahams, Reiss & Sharpe (2013) as "an overarching term that refers to any type of science teaching and learning activity in which students, working either individually or in small groups, are involved, as an important element of what they are doing, in manipulating and/or observing real objects". This can be extended to cover a wide range of activities covered by 'Working Scientifically' and extended beyond the manipulation of equipment to "the promotion of higher order thinking skills, such as problem-solving and scientific thinking." (El Masri, Erduran & Ioannidou, 2021).

The purposes of practical work are widely discussed in scientific education literature (Cramman et al, 2019; Moore et al, 2020; Score, 2009; Wilson, Wade & Evans, 2016). The Gatsby (2017) 'Good Practical Science' report identifies five key purposes of practical science in schools through a review of the literature and an international survey. These are an appropriate summary for the purpose of this project and are:

A To teach the principles of scientific inquiry
B To improve understanding of theory through practical experience
C To teach specific practical skills, such as measurement and observation, that be useful in future study or employment

D To motivate and engage students

E To develop higher level skills and attributes such as communication, teamwork, and perseverance.

The assessment of practical work in science

The assessment of practical work in science is much debated and various approaches have been used for this assessment in English national qualifications (Evans, 2013). This is because there is not the same level of agreement or acceptance about how this assessment, in the context of a national examination system, should be carried out, as there is about its value in the teaching and learning of science. There is variation around the world about how the assessment of practical science contributes to high-stakes assessments that are part of national examination systems. For example, in Singapore, 16 year old students take an end-of-course practical examination as part of their O level science qualification and in New Zealand, 16 year olds write a practical report which is assessed by their teacher and moderated within the science department (Erduran et al, 2020). This variation may well lead to the accepted pedagogical benefits and the purposes of practical science being undermined when trying to meet the requirements of a national examination system (Wilson, 2013).

In addition, the requirements of high-stakes assessments, those where there is a significant impact on any or all of students, teachers, and schools (Childs & Baird, 2020), are known to play a role in determining science curriculum content and the approaches taken by teachers when planning and delivering practical lessons (Erduran et al, 2020). Indeed, if assessments are carried out internally in schools and colleges by teachers, it is

recognised that accountability measures, upon which schools and teachers are judged, influence the way in which these assessments are managed (Ofqual, 2013a). In addition, the accountability system promotes a situation where there is a lack of trust (Baird & Elliott, 2018).

In England, the assessment of practical work in science has historically been carried out in various ways in the context of a national examination system. Students sit high-stakes qualifications at age 16 in a range of core subjects including English, mathematics, and science as well as in optional subjects such as humanities, modern foreign languages, and arts. These qualifications, called GCSEs (General Certificate in Secondary Education), are offered by four independent exam boards. Schools and colleges can choose to enter their students for qualifications from any exam board and so it is a competitive but regulated market (Opposs, 2016). GCSEs were first awarded in 1988 and have been through various reforms in the past 30 years which, in the case of GCSEs in science, have resulted in different ways of assessing practical science.

After sitting GCSE qualifications at age 16, students have a choice of post-16 qualifications including in vocational and technical subjects such as sports science, construction, and digital production. Students can also choose a more academic pathway in the form of GCE (General Certificate of Education) A level qualifications which are high-stakes assessments taken by students usually at age 18. These qualifications have been part of the post-16 qualification landscape since 1951, are taken on a subject-by-subject basis and are primarily used for university admission. These, too, have been reformed a number of times over the last 70 years and again, the assessment of practical skills in the science A levels has taken a range of approaches.

The assessment of practical work in the A level sciences in England

Approaches to the assessment of practical work in the A level sciences have included a formal practical examination, used as an assessment method from the first A level syllabi through several iterations up to 2010. Students were required to carry out a selection of experiments, report on the results and answer questions in examination conditions under the supervision of an invigilator. A level science qualifications awarded between 2010 and 2016 used coursework to assess practical work. This was an internal assessment, taken in school that was overseen and marked by teachers. Coursework tasks were moderated by the exam board and marks contributed towards the final A level grade (Baird et al, 2019). Coursework, involving teacher assessment, was viewed as an essential part of science qualifications, assessing skills such as the manipulation of instruments and taking measurements – skills that cannot be easily assessed in a written examination (Torrance, 2018).

The current approach to the assessment of the practical skills of A level science students was developed in time for the most recent reform of all A level qualifications, the first of which were awarded in 2017. These reforms began with the 2010 White Paper entitled "The Importance of Teaching" (Department for Education, 2010) in which the government made clear its intention to reform GCSE and A level qualifications. In the case of A levels, there was concern about the unintended consequences of a modular structure on students' preparedness for university (Department for Education, 2012). These consequences included the ability to re-sit modules all through the course so approaching examinations with the view that there was always another chance (Ipsos

MORI, 2012) as well as students' ability to make connections between different parts of the subject content. Reform was proposed to address these concerns.

In addition, there was widespread discussion and research had been carried out into potential issues with coursework in both GCSEs and A levels, including in the science subjects. By the year 2000, concerns were expressed about varying teacher and parental support, the ability to authenticate work as being done entirely by students, marking reliability and the possible impact of coursework on grade inflation (Opposs, 2016). Teachers did not regard the use of coursework in science as having much educational merit as the approach had reduced the variety of practical activities to only a few so as to ensure that students could access the full range of coursework marks (QCA, 2005).

There was also concern about the functioning of the assessments and the marks students achieved when compared to marks in the written examinations (Opposs, 2016). This can be seen in Figure 1 below which shows the mark distribution for the coursework component of a physics A level qualification (left) compared with that for a written examination taken by the same cohort (right). The coursework assessment does not differentiate well between students as their marks are negatively skewed and bunched at the top end of the mark distribution (Wade & Abrahams, 2015).

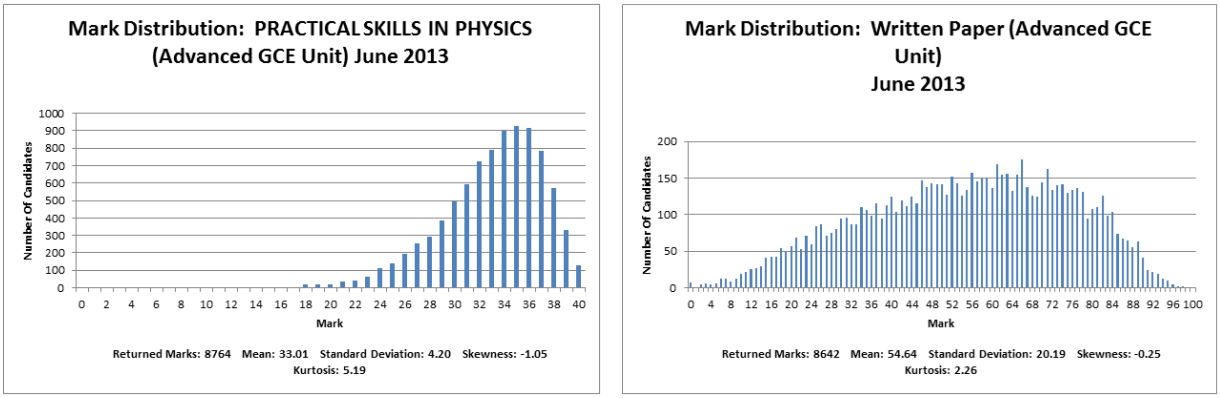


Figure 1. Variation in mark distribution for practical and written assessment

These concerns about coursework were articulated by Ofqual, the examinations regulator in England, in their consultation seeking views about a new method of assessing practical work in science A level qualifications (Ofqual, 2013b). This method, crucially, proposed separating the direct assessment of practical skills (Abrahams et al, 2013) from the overall A level grade. The science community, including employers and higher education institutions, hold the opinion that the development of dexterity, manipulation of apparatus and other laboratory work are of high value in the workplace as well as for further study (Abrahams et al, 2013). The new method proposed that a student would be awarded an endorsement alongside the A level grade indicating whether they had reached the required competency in a range of subject specific practical skills. These skills would be demonstrated by the completion of at least 12 practical activities over the 2-year A level course. Students would also be indirectly assessed on their practical skills (Abrahams et al, 2013) as it was proposed that at least 15% of the marks in the A level written examinations would come from questions based on practical skills. This approach aimed to embed the practical work throughout the A level science course (Cramman et al, 2019).

It is not an understatement to say that this proposal caused consternation amongst the science community who expressed the views that this approach devalued the importance of practical work and would lead to students developing far fewer practical skills (Carter, 2014). In spite of this opposition and with assertions that practical science remained an important part of science qualifications (Ofqual 2013b), from 2015, A level practical work assessment changed and is now made up of two parts. Written examinations have at least 15% of marks assigned to questions that indirectly assess practical skills and the Practical Endorsement is awarded as a 'pass' alongside the main A level grade if students have shown competence in a range of subject specific skills that have been developed throughout the 2-year course.

The intention of these reforms to the assessment of A level science practical skills was to ensure that a wider set of practical skills were developed by students. Criticisms about the previous approach included the requirement of 'jumping through hoops' to achieve as many marks as possible and the narrowing of the practical opportunities for students to a very small range of activities to achieve the intentions of the qualification (QCA, 2005). A pass result in the endorsement was to be a fair reflection of the competency of the student in their science practical skills and was to be a clear indication to end users of these results of this competency of the student.

The exam boards worked collaboratively to define these skills alongside criteria that teachers use to assess their students – the Common Practical Assessment Criteria (CPAC) (Ofqual, 2016 pp 15-16). The CPAC require students to develop skills in following written procedures, applying investigative approaches and methods, safely using a range of practical equipment and materials, making and recording observations and research,

referencing and reporting (Ofqual, 2016 pp 15-16). The A level science specifications, developed and implemented by the exam boards, provide further instructions as to the practical activities that need to be carried out – to allow teachers to assess students in the CPAC and to ensure students are prepared to answer the questions in the examination papers that indirectly assess practical skills.

To support teachers in their assessment using the CPAC and to ensure the standard of this assessment, exam boards use a monitoring process that involves visits to schools and colleges to observe practical work being carried out and checks on the record keeping for the endorsement (Ofqual, 2016 pp 16-17).

Research into the impact of this qualification reform on the practical skills of A level science students looked at the practical skills of first year undergraduates from a range of Higher Education Institutions, some of whom had studied their science A level prior to the reform, and some had done so after the reform (Ofqual, 2019). Overall, it was concluded that the practical skills of post-reform chemistry and physics undergraduates had not declined and, in fact, the skills of post-reform biology undergraduates may have increased. In addition, Opposs (2016) reports that "In most schools, being freed from the need to carry out set piece practicals with the aim of maximising marks has allowed teachers to provide students with much better opportunities to develop their skills and for the exercise and development of these skills to be better integrated into teaching and learning over the full course of study."

The impact of the COVID-19 pandemic on practical work in schools and colleges

On 18th March 2020, schools and colleges in England closed to most students in response to the COVID-19 pandemic (Department for Education, 2020). In addition, GCSE, AS and A level examinations were cancelled. Disruption to education has continued to the present time with public examinations also being cancelled in 2021. Examinations went ahead in summer 2022 but it was recognised that students had faced disruption to their education as a result of the pandemic, so measures were put in place to support students in their revision and in the examinations (Department for Education, 2021).

This disruption to education had an impact on the practical work that was carried out in schools and colleges during the pandemic. A level science teachers were encouraged to prioritise opportunities for 'hands on' practical work to their students to provide evidence for the award of the Practical Endorsement (Ofqual, 2021a). However, school closures and online learning may have resulted in fewer opportunities for students to develop practical skills in comparison to prior to the pandemic. Public health restrictions required schools to operate within 'bubbles' and this meant that laboratory space was not accessible to all students. The cleaning of shared practical equipment and the time needed between uses made practical work a challenge in many settings. Schools and colleges were supported by CLEAPSS, an organisation which provides services such as advice about health and safety and risk assessments to science teachers, to carry out practical work safely in the context of the public health restrictions[1]. However, it is still possible that students carried out fewer practical activities and spent less time each week on

[1] Information can be found on the CLEAPSS website here https://science.cleapss.org.uk/resource-info/gl343-guide-to-doing-practical-work-during-the-covid-19-pandemic-science.aspx

practical work than in a normal year. With the possibility of fewer opportunities for practical work, there may have been different types of practical work carried out e.g., more that allowed assessment of following written procedures than those that allowed students to develop investigative skills.

In light of these possible restrictions on practical work and following public consultation, Ofqual relaxed the requirements for the Practical Endorsement for students being awarded the endorsement in 2021 (Ofqual, 2020a) and in 2022 (Ofqual, 2021a). Science teachers were allowed to use fewer than the previous requirement of 12 practical activities to assess the CPAC as long as students were still able to demonstrate their competence in all of the criteria. This means that it is possible that the Practical Endorsement may have been awarded on the basis of a smaller number of practical activities than in a usual year. It also means that some students may have been awarded the endorsement on the basis of far fewer practical activities than others. There is a range of evidence about the impact of the pandemic on teaching and learning but there is no evidence about the extent of the impact of the pandemic on teaching and learning of practical science skills (Ofqual, 2021b). This variation in the amount of practical work needed for assessment and the lack of evidence of the impact on practical science skills raise questions of fairness.

Conceptions of fairness

The impact of the pandemic on the national qualifications system has raised issues of fairness to the forefront of the minds of the public and those that are involved with the design and development of assessment. Fairness has been raised more frequently in academic writing about educational assessment in recent years than previously to such an extent that some authors have suggested that it should have the same status as validity and

reliability (Nisbet & Shaw, 2020 pp 2-3). The increasing importance being placed on fairness began before the pandemic and has grown in the last two years. The authors go on to say that in many cases, it is not necessarily clear what is meant by the term 'fairness' and that it is possible to view fairness in a variety of ways i.e., through different 'lenses'. Nisbet & Shaw (2022) add to these thoughts by emphasising that "Fairness should be perceived as being located on a continuum" – it is not a binary idea where assessment is either fair or unfair but that there can be degrees of fairness.

It is important to establish what is meant in the context of this research by fairness and how it can be applied to the assessment of practical work in the A level sciences as a result of the pandemic. Fairness has been defined many times in the context of educational assessment but there is no agreed definition of it as a technical term (AERA et al, 2014 p. 49). There is, however, the definition of "the quality of treating people equally or in a way that is right or reasonable" in the Cambridge English Dictionary (2022). This is an appropriate place to start when considering the concept in the context of educational assessment. Indeed, it is included as part of first of four views of fairness in the most recent edition of The Standards for Educational and Psychological Testing, referred to from now on as the *Standards* (AERA et al, 2014 p. 51-4):

1. Fairness in treatment during the testing process – this includes the expectation that tests assess students on the intended construct and that approaches to scoring and the conditions of test-taking are the same for all. In order that all students are able to demonstrate their ability on the test construct, adjustments may be needed to address the specific needs of some e.g., a test provided in a large font for a student with a visual impairment.

2. Fairness as a lack of measurement bias – the test needs to avoid bias in measurement across identifiable sub-groups within a cohort due to features such as sex, disability, race, and ethnicity.
3. Fairness in access to the construct(s) being measured – when individual characteristics of a student do not impact their ability to demonstrate their knowledge and understanding of the measured construct. It is important to provide access to a test construct whatever the characteristics of the student. For example, if a test contains specific language unrelated to the test construct, a student who is not a native English speaker may be unable to access the construct of the test.
4. Fairness of individual test score interpretations for the intended uses – an interpretation of a test score needs to take account of a students' individual characteristics and their interrelationship with the context of the test.

The first view includes the idea that it can be fair to treat people differently if their individual characteristics warrant this different treatment. This is equitable as opposed to equal treatment (Nisbet & Shaw, 2019).

It is beyond the scope of this research to enter into a detailed consideration of how fairness is related to and possibly conditional for reliability and validity, but it should be said that there is a wealth of literature that discusses the relationship of these three important concepts in educational assessment (AERA et al, 2014; Camilli, 2006; Kane, 2010; Nisbet & Shaw, 2020; Shaw & Nisbet, 2021; Sireci & Randall, 2021; Worrell, 2016; Xi, 2010; Zieky, 2016). There is no doubt that there is some agreement with Zieky (2016) in his view that "fairness is a matter of validity" and that a fair test is also one that is valid. Xi (2016) says that fairness is a part of validity and that if fairness is undermined,

then this also compromises the validity of a test. The *Standards* (AERA et al, 2014 pp 49-62) uses the term validity all through the chapter that describes the role of fairness in testing indicating the link between the concepts. Caines, Bridglall & Chatterji (2014 p. 12) make the point that "Validity is a complex and evolving construct, especially when considering issues of fairness."

Nisbet & Shaw (2020, pp 25-28) discuss the relationship between fairness, validity, and reliability. They consider whether fairness and validity are linked such that any test that is valid is also fair or whether any test that is fair is also valid. They then consider whether a test must be fair to be considered valid and whether a test must be valid to be considered fair. There is also the possibility that validity and fairness are not related. Nisbet & Shaw conclude that these ideas are important to educational assessment, and, in their view, fairness is an essential requirement for validity. Their discussion of the relationship between fairness and reliability links the latter to test scores that are stable over time, free from bias and repeatable in terms of resulting in precise scores. They determine that an unreliable test would also be unfair but that the inverse is less easy to determine. Reliability and fairness are, in their opinion, related but separate (Nisbet & Shaw, 2020, p 28).

Worrell (2016, pp 283-285) describes validity and reliability as the two leading concepts in educational assessment. He goes on to say that fairness has become increasingly prominent in each subsequent edition of the *Standards* since it was first mentioned in the 1974 edition. The latest 2014 edition gives fairness its own chapter and suggests that validity, reliability, and fairness should be described as the 'big three' of educational assessment. At the start of the fairness chapter, the *Standards* includes validity in the

comment that “Fairness is a fundamental validity issue and requires attention throughout all stages of test development and use.” (AERA et al, 2014).

The progressively higher prominence of fairness in subsequent editions of the Standards indicates that there has been a gradual improvement in the understanding of how fairness is important to educational assessment over the last 50 years (Sireci & Randall, 2021). This improvement in understanding is also shown by Nisbet & Shaw (2020) in their book ‘Is Assessment Fair?’.

The following distinctions between a range of senses of fairness are taken from the above book (Nisbet & Shaw, 2020). They are relevant to educational assessment and to the consideration of the impact of the pandemic on the direct assessment of practical skills in the A level sciences.

1. Formal fairness – the correct application of a rule or design with a focus on the application of the rule rather than any discussion of the appropriateness of the rule.
2. Implied contractual fairness – an assessment is fair if it is what those affected are legitimately expecting. If an assessment does not meet expectations, students and teachers may feel that it has been unfair.
3. Relational fairness – the treatment of relevantly like cases alike.
4. Retributive fairness – when the outcome of an assessment is appropriate in comparison to what has gone before and so is deserved.
5. Consequential fairness – it should not be possible for outcomes of assessment to be used for unfair actions in the future.

6. Retrospective fairness – an assessment may be unfair if it is the result of unfair actions in the past.

Fairness of A Level practical science assessment during the pandemic

In terms of possible threats to the fairness of the direct assessment of practical work in the A level sciences as a result of the pandemic, the first and fourth views from The Standards for Educational and Psychological Testing (AERA et al, 2014 p. 51-4) have the most relevance. These are fairness in treatment during the testing process and fairness of individual test score interpretations for the intended uses respectively.

It is likely that not all A level students were treated the same as a result of the pandemic. Allowing teachers to assess students on the CPAC using fewer than the normal 12 or more practical activities may have resulted in construct underrepresentation (Shaw & Imam, 2013) – some students may go onto inappropriate higher education courses or employment which may be a matter of fairness. This might be particularly the case if different types of practical activity were used to assess students to those used prior to the pandemic. There may not be fairness in the interpretations of the result of the Practical Endorsement as it is not possible to identify from a 'pass' result whether students have been assessed on fewer than the normal requirement of 12 practical activities. This may be unfairness between students within the 2021 or 2022 cohort or between them and students in different cohorts. It may be that the endorsement result does not have the same meaning for all students across years of assessment, a concept described about assessments more generally by Camilli (2006).

However, there are other conceptions of fairness at play in this situation. It may not be considered fair to expect students to be assessed on all 12 practical activities in their individual context of the impact of the pandemic; for many, this may have simply been impossible and through no fault of their own. Students may have been unwell themselves and absent from school or college or absent due to the need to care for others and this absence may have affected how much practical work was completed and assessed. As a result of this, it may not have been fair to try to treat students within a cohort in the same way as some may not have had any absence at all. The impact of the pandemic on students and their practical assessment may have varied according to characteristics protected by the Equality Act (2010) and students' special educational needs and disability (SEND). The situation of the pandemic may have been particularly unfair to students who are already vulnerable.

There are clearly different conceptions of fairness which can be brought to bear in the context of the assessment of practical science skills during the pandemic. Nisbet and Shaw's approach to fairness is helpful in allowing us to tease them apart. In this context, there are following definitions of fairness within their typology:

1. Formal fairness – it may not have always been possible to keep to the rules of the assessment of the CPAC, even though Ofqual revised these rules to allow assessment using fewer than the usual 12 required practical activities, due to public health restrictions and closure of schools and colleges to most students.
2. Implied contractual fairness – because of the pandemic, the CPAC was assessed in a way that was not expected by students and teachers.

3. Relational fairness – students being assessed in the CPAC may not have been able to carry out the same number and type of practical activities as students in their school and college, other students taking the assessment in the same year or previous or future cohorts of students. The pandemic may have had an impact on a students' opportunity to learn (Camilli, 2006) in the sense that they may not have had access to the range of practical work opportunities that other students may have had. Relational fairness was impacted in both the assessment of the CPAC and the learning that students could access related to practical work.
4. Retributive fairness – a 'pass' result in the Practical Endorsement may not be a deserved result for students as a result of the pandemic. They may have completed less or different practical work to that done in normal circumstances.
5. Consequential fairness – the pandemic may have had the impact that a students' practical skills are weaker than a 'pass' result in the Practical Endorsement would suggest. The negative consequences of this are that students may be able to access higher education courses or employment in situations such as laboratories on the basis of their pass, but their practical skills may not be as good as would be needed either for undergraduate study or to carry out the requirements of a job.
6. Retrospective fairness – an unfair action in the past may have been to have ever decided to assess practical work in science in the first place which may now make the current method of assessment unfair.

Nisbet & Shaw go on to say that it is important to clarify who the (un)fairness affects. It is common to assume that this would, in the context of this research, just be the student who is awarded the Practical Endorsement. However, if a student uses their 'pass' result to access a university course and this pass indicates weaker practical skills than it did in the

years prior to the pandemic, then the university administration may also consider this to be unfair. In addition, the lecturers of undergraduate practical courses may well have to provide additional support for students who do not have sufficiently strong practical skills to cope with practical work at this level which may be unfair. This may also be the case for laboratory staff employing students with weaker practical skills than indicated by a 'pass' in the past.

It is reasonable to think that the respondents to the survey, teachers and monitors, may view relational fairness to have had the most impact on the assessment of practical skills as a result of the pandemic. It is likely that there was variation in the disruption that students experienced to their education in 2020-21 and again in 2021-22 due to a long list of factors including those discussed in the review carried out by Ofqual into learning during the pandemic (2021b). One of the general conclusions of this review was that "even students from very similar backgrounds, registered in the same class at the same school, might have experienced quite different levels of learning loss, depending on the particular circumstances of their learning during the pandemic". This is particularly appropriate to a discussion of the impact of relational fairness in the context of A level science practical skills assessment – if these students may have had very different experiences, it is likely that students in different schools and colleges, locations and of different backgrounds may well have been treated differently.

In this research, the typology provides a tool with which to investigate participants views of the fairness of the assessment. The impact of the pandemic provides an opportunity to test the validity of the typology.

Research questions and hypotheses

The aim of this research is to investigate the impact of the COVID-19 pandemic on the direct assessment of practical work in the A level sciences, and in particular on the fairness of the assessment. This is because a pass result in the endorsement in 2017-2019 may not mean the same as in 2020-2022 as the rules were different and students were able to be assessed on the CPAC having completed less practical work. This work will, in addition, test the validity of Nisbet and Shaw's typology of fairness.

The following research questions are investigated:

1. How has the pandemic impacted the delivery of hands-on practical work to A level science students?
    a. Was there a reduction in the hours of practical work completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic?

    b. Have students from the 2021 cohort completed fewer hours of practical work per week than students in the 2022 cohort?

    c. Did the number of hours of practical work completed by students per week vary by school type before the pandemic, in 2021 and in 2022?

    d. Was there more of a focus on the 12 required / core practicals in 2020-2021 than before the pandemic and 2021-2022?

e. How has the type of practical work completed by students in the 2021 and 2022 cohorts changed in comparison to before the pandemic?

2. How has the pandemic impacted the fairness of assessment of practical skills in the A level sciences?
    a. Which senses of fairness do respondents believe have been most impacted by the pandemic?

3. How does the number of hours of practical work completed by students impact on respondents' perceptions of the ability of the teacher to judge fairly their competence against the CPAC?
    a. What is the relationship between the number of hours of practical work per week and respondents' views of different senses of fairness?

    b. Do teachers / monitors think that relational fairness is the sense that had or will have the most impact as students completed fewer hours of practical work in 2020-2021 than before the pandemic?

    c. Do teachers / monitors think that formal fairness will have been impacted as the rules for the assessment of the CPAC have been relaxed during the pandemic?

4. How well is Nisbet and Shaw's typology of fairness measured by the interviews and survey?

a. Do the survey items that ask about the same senses of fairness correlate together and form distinct factors?

b. Do the different senses of fairness have different relationships with respondents' views of the impact of the pandemic?

The following alternative hypotheses are investigated:

- There will be a reduction in the hours of practical work completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic.

- Students from independent schools and colleges will have completed more hours of practical work per week before the pandemic, in 2021 and in 2022 in comparison to students from state funded schools and colleges.

- The type of practical work completed by students will vary for the 2021 and 2022 cohorts in comparison to before the pandemic.

- Responses on items will correlate together and form scales with high internal reliability.

- Teachers / monitors will have the view that relational fairness is the sense of the fairness that has or will have the most impact. A pass result in the endorsement in 2017-2019 cannot be said to mean the same as in 2020-2022 as the rules were

different and students were able to be assessed on the CPAC having completed less practical work.

- Teachers / monitors will have the view that formal fairness will have had an impact on the results of the endorsement. The rules for the assessment were flexible during the pandemic, so it is hard to judge what competence in the CPAC really means from one student or school and college to another.

- The typology to describe fairness in assessment is exhaustive and mutually exclusive.

Method

The impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences was investigated using a mixed methods approach. This approach was chosen to allow a more comprehensive understanding about fairness in this context than a single method would have provided. The participants of the research are described below along with the ethical issues considered. The approach to the structure and design of the survey is described followed by the approach to the interviews and the way that questions were identified. An outline of the type of data available from the survey is given followed by a brief description of the approach to the analysis of the survey data.

Survey

Primary data were collected for this research by surveying a sample of teachers and exam board monitors and those that perform both roles. The teachers were those of the A level science qualifications, namely biology, chemistry, physics, and geology. These subjects were chosen as they have the same method of assessment of practical skills – the Common Practical Assessment Criteria (CPAC) – used to award the Practical Endorsement, a result of the assessment of practical skills that is awarded separately to the main A level grade.

To be eligible to complete the survey, teachers had to have experience of using the CPAC with students both prior to and during the COVID-19 pandemic. Exam board monitors

had to have experience of monitoring the practical work in schools and colleges over the same period. Exam board monitors are either recently retired or current teachers.

The survey was published online as this was convenient to potential respondents, likely to reach a wide population of teachers and exam board monitors and overcome the researcher's geographical separation from the population of potential respondents. It also allowed design flexibility so that teachers and exam board monitors and respondents who carry out both roles could all access the same survey and take appropriate routes through the questions. This method meant that respondents could answer anonymously if they wished.

The survey was advertised to potential respondents using contacts established by the researcher as a result of prior employment as a science teacher and current employment with Ofqual. This included representatives from the science teams of the exam boards who meet regularly to oversee the monitoring of the Practical Endorsement. The researcher was also able to share the survey with relevant contacts at the learned societies for them to circulate to their members. This included the Royal Society, the Royal Society of Biology, the Royal Society of Chemistry, the Institute of Physics, and the Association of Science Education. The survey was shared widely on social media, including Twitter and Facebook, with the researcher requesting help with sharing the survey as widely as possible from people from science education active on social media.

Respondents completed the survey voluntarily and 81 surveys were submitted in total. It was not possible to determine the number of A level science teachers and monitors currently performing these roles. A representative sample would be the percentage of the population of these groups necessary to replicate the opinions of fairness in the context of

the A level science practical skills assessment and the pandemic and so it was determined that the 81 respondents had raised views that could be considered to be representative.

The respondents were 54 teachers, 16 monitors and 11 who carry out both roles. Of the 65 teacher respondents, 43 were from state funded schools and colleges and 22 were from independent schools and colleges. 22 biology teachers completed the survey, 28 were chemistry teachers and 15 were physics teachers. No geology teachers responded. Average cohort size ranged from 3 students to 720. Experience of A level teaching ranged from 3 to 42 years.

Of the 27 monitor respondents, 6 monitor biology, 10 monitor chemistry and 11 monitor physics. No geology monitors completed the survey. Experience of monitoring ranged from 3 to 7 years. 17 monitors worked for the exam board AQA, 4 worked for OCR, 5 worked for Pearson and 1 worked for WJEC-Eduqas.

Interviews

Respondents to the survey were invited to leave their name and email address at the end of the survey if they were interested in exploring their views about fairness of the assessment of practical work in the context of the pandemic further by participating in an interview via Zoom. This self-selected group of 28 from a total of 81 (35%) were invited to be interviewed and 11 responded with their availability in the required time frame. This group of 11 was made up of 4 chemistry teachers, 1 physics teacher, 1 chemistry teacher and monitor, 2 chemistry monitors, 2 physics monitors and 1 biology monitor. This is quite representative of those who completed the survey, but it is noted that no biology

teachers were interviewed. The interviews were carried out using a semi-structured approach. This was determined to be the most appropriate method for data collection to allow the interviewer to prompt the participant with open-ended questions depending on their responses. This meant that the wording and sequence of the questions were specific to the participant (Cohen, Manion & Morrison, 2018 p.511).

Participants were interviewed for 30 minutes, and the discussion began with an open-ended question asking the participant to describe their experience of the delivery and / or observation of practical skills in the A level sciences during the pandemic. Questioning then continued to explore the concepts of fairness as described above.

Ethical issues

The CUREC application approval for this research is in Appendix F. Consideration was given to the data collected using the survey and interviews. It was important that teachers and monitors were given the option to answer the survey anonymously if they wished. The survey was constructed so that respondents were not identifiable unless they provided their name and email address as an expression of interest to be interviewed. Informed consent was obtained by asking participants to read information about the research and agree to their participation at the start of the online survey and, in addition, at the start of the interview process.

Participants were advised that they could withdraw from the survey at any time prior to submitting their answers. They were also advised that if they did not provide contact details, once answers had been submitted, their responses would remain anonymous, but

withdrawal would no longer be possible. If participants did provide their contact details, it was possible to identify their responses after submission, so participants were able to contact the researcher to request withdrawal at any point up to submission of the project. Participants could withdraw from the interview stage of the research by contacting the researcher at any point up to submission of the project.

All survey and interview data were stored within the online survey tool or on a Nexus 365 OneDrive. A Zoom transcription was generated at the end of each interview so that it was not possible to identify participants beyond knowledge of their name and email address. This information was deleted at the end of the data collection phase and interviewers identified below by role and number e.g., teacher 1, exam board monitor 2.

Survey and interview questions

Survey questions for teachers, exam board monitors and those that had both roles are in Appendix G. They were divided into two main sections. Routing through the survey allowed the researcher to ask questions about the nature and frequency of practical work opportunities either provided to students or observed during monitoring visits to each group separately. Following this, all respondents answered the same questions on the senses of fairness. These questions were based on each sense of fairness and tailored to the context of the assessment of practical work in the context being researched. An example of an item for each sense of fairness is below:

- Formal: the rules in place for the assessment of the CPAC were not always appropriately followed.

- Implied contractual: the assessment of the CPAC was not carried out as students and teachers had legitimately expected.
- Relational: not all students were treated the same way in the assessment of the CPAC.
- Retributive: it was not always possible to appropriately reward students for their competence in their practical skills.
- Consequential: students will be treated unjustly in the future based on their result in the Practical Endorsement.
- Retrospective: the assessment of the CPAC has never been a fair assessment method of practical skills.

The questions were not presented to the respondent in groups according to each sense of fairness but mixed up to try to avoid respondents drawing their own conclusions about each sense of fairness.

Interview questions are in Appendix H and were initially determined using the responses of the participants to the survey questions. The researcher wanted to use these as a starting point to any detailed discussion about fairness, so the final list of questions was determined once the survey closed. The questions focused on the different senses of fairness that make up the typology being evaluated in this research and how they applied when using the CPAC both before and as a result of the pandemic, right up to the ways in which A level science practical skills were being assessed this academic year (2022).

Analysis of survey and interview data

To answer the questions at the core of this research, two strands of enquiry emerged. The first was interest in the impact of the pandemic on practical work for A level science students. Survey and interview questions focused on what it had been like as an A level science teacher and / or monitor over recent years, how much practical work they had been able to carry out with students and how it had been different to that done in the years before the pandemic. The second strand focused on the fairness of the ways in which assessment has been carried out as a result of any public health restrictions put in place and partial school closures that occurred during the pandemic.

The first strand resulted in a five-part research question and all parts of this were answered using a series of statistical measures to evaluate the impact of the pandemic on the delivery of hands-on practical work to A level science students. To find out if the pandemic had resulted in fewer hours of practical work being delivered to students, statistical analysis of survey responses from teachers was conducted. A one-way repeated measures ANOVA looked at whether there was any difference in the number of hours of practical work carried out per week at three time points – before the pandemic, in the academic year 2020-21 and then in the academic year 2021-22. This analysis was then enriched with comments derived from content analysis of the interviews and free responses to survey questions based on the impact of the pandemic on how much practical work had been possible in schools and colleges.

To assess whether the number of hours of practical work completed per week over the course of the pandemic varied by school type, a two-way mixed ANOVA was conducted.

Teachers had identified their school type as part of the survey so this information was collated to form two variables – independent school or college and state funded school or college. The latter included academies, maintained school or colleges, further education colleges, sixth form colleges and faith schools.

As described above, one of the features of the Practical Endorsement is the requirement that A level science students complete at least 12 practical activities. These are based on a list of practical skills determined by the Subject Content for each A level science subject but the actual approach to the completion of at least 12 practical activities is determined by the specification chosen by the teacher. The exam board AQA uses an approach of 'required practicals', Eduqas uses 'specified practical work', OCR requires students to complete an activity from each of 12 'practical activity groups', known as PAGs and Pearson uses 'core practicals'[2]. Within the context of this terminology, teacher and monitor respondents were asked in the survey to what extent the focus was on just completing these 12 practical activities at the same three time points: before the pandemic, in the academic year 2020-21 and then in the academic year 2021-22. A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in the focus of practical work at these three time points. This analysis was then enriched with comments derived from content analysis of the interviews and free responses to survey questions about any change in focus.

Assessment using the CPAC requires a range of practical skills to be developed by students. For example, the first competency requires students to follow written procedures

[2] From this point, these different approaches will be grouped together and referred to as 'required practical activities'.

and the second requires some element of investigative work. To determine any impact of the pandemic on teachers' ability to conduct practical activities of different types, survey respondents provided information about the types of activity carried out at the same three time points: before the pandemic, in the academic year 2020-21 and then in the academic year 2021-22. Respondents were able to indicate any that were delivered from a list of:

- Experiments following a method
- Experiments involving some element of investigation
- Whole open-ended investigations
- Teacher demonstrations
- Simulations
- Fieldwork
- Other (the next question asked for details about this type)

Chi-squared tests were conducted to determine whether there was a significant difference in the types of practical work completed by students before the pandemic compared with students in the 2021 and 2022 cohorts.

The remaining 3 research questions were based on the second strand of enquiry: the impact of any differences in approach on the fairness of the assessment of practical work in the A level sciences. The survey contained 26 items based on the Nisbet & Shaw fairness typology described above, with items designed to ask about the 5 senses of fairness being evaluated in the context of the impact of the pandemic on the assessment of A level science practical skills. Respondents were asked their opinion using a 5-point Likert scale with the options of strongly agree, agree, neither agree or disagree, disagree

and strongly disagree. Scales were created of these data to the 26 fairness items, grouping them according to the sense of fairness that was asked about – formal, implied contractual, relational, retributive and consequential. A scale was also created for all 26 items. To create the scales, the Likert scale scores for each item within a group representing a sense of fairness were added together and divided by the number of items in that sense of fairness so that the scale back was still expressed as the clear 0-5 scale.

A Cronbach's Alpha was calculated to measure the internal consistency of the groups of items designed to measure different senses of fairness and for all of the items overall. This provided information as to whether the items were measuring a common construct.

A one-way repeated measures ANOVA was conducted to investigate whether there was any difference in the impact of the pandemic on the different senses of fairness. This analysis was then enriched with comments derived from content analysis of the interviews and free responses to survey questions based on the impact of the pandemic on fairness.

Pearson's correlations were conducted to determine any relationship between hours of practical work at the three time points and respondents' views of the impact of the pandemic on different senses of fairness. This analysis was then enriched with comments derived from content analysis of the interviews and free responses to survey questions about the any relationship between the number of hours of practical work per week and the different senses of fairness and the relative importance of relational or formal fairness.

The final research question considered how well Nisbet and Shaw's typology of fairness was measured by the interviews and survey to evaluate the ability of the typology to describe fairness in assessment in an exhaustive and mutually exclusive way. An exploratory factor analysis was conducted to determine whether the fairness items that asked about the same senses of fairness correlate together and form distinct factors. Open survey items and responses to interview questions were coded and analysed to investigate respondents' views on the effect of the senses of fairness on practical skills as a result of the pandemic as well as more broadly about the impact of the pandemic on the provision of practical opportunities for students.

Results

Open survey items and responses to interview questions were coded using NVivo and analysed to investigate respondents' views on the effect of the senses of fairness on practical skills as a result of the pandemic. Related material was collected into containers within the software called nodes to identify themes in the comments. Each sense of fairness was a node and comments about the impact of the pandemic on practical work, the amount of practical work that was possible and that the approach was as fair as it could be in the circumstances were also identified as nodes. Table 1 below shows the frequency that each node was identified in the comments.

Table 1. Coding of survey open items and interview responses

Node	Frequency of references
Overall fairness	16
Formal fairness	21
Implied contractual fairness	6
Relational fairness	78
Retributive fairness	21
Consequential fairness	2
Retrospective fairness	11
As fair as it could be in the circumstances	20
Impacts on practical work	30
Amount of practical work	33

The number of hours of practical work per week

Questions in the survey asked teacher respondents how many hours of practical work students completed per week before the pandemic, in 2020-21 and in 2021-22. The hypothesis is that there will be a significant reduction in the hours of practical work

completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic.

Table 2. Descriptive statistics of hours of practical work per week before the pandemic, in 2020-21 and in 2021-22

Descriptive statistics	Hours of practical work completed each week		
	Pre-pandemic	2020-21	2021-22
N	64	64	62
Mean	1.49	0.81	1.47
Std. Deviation	0.74	0.57	0.74
Range	3	3	3
Minimum	1	0	1
Maximum	4	3	3

Data in Table 2 shows that the mean number of hours of practical work per week was lower in 2020-21 than pre-pandemic and in 2021-22. Table 2 also shows that the mean number of hours of practical work per week in 2021-22 was similar to that carried out before the pandemic.

A one-way repeated measures ANOVA was conducted to determine whether there was a significant difference in the number of hours of practical work pre-pandemic, in 2020-21 and 2021-22. There was one outlier in the data for 2020-21 as assessed by boxplot, as seen in Appendix A, and the data did show kurtosis at the 2021-22 time point, as assessed by calculation of the kurtosis (the sharp peak was caused by the most common value for the hours of practical work being 1). The one-way repeated measures ANOVA was conducted with and without the outlier. There was little material difference in the outcome as seen in Appendix A, and so the analysis was carried out using the full data set.

The assumption of sphericity was not met, as assessed by Mauchly's test of sphericity, χ^2 (2) = 34.65, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied (ε = 0.70). The pandemic resulted in a significant reduction in the number of hours of practical work in 2020-21 (F(2,85) = 53.98, p = <0.001).

Post hoc analysis with a Bonferroni adjustment revealed that there was a reduction in the number of hours of practical work per week in 2020-21 (M = 0.81, SD = 0.60 hours) compared to the pre-pandemic number of hours (M = 1.50, SD = 0.75 hours), a significant mean reduction of 0.69 hours, 95% CI [0.27, 0.49], p = <0.001. There was also a reduction in the number of hours of practical work per week in 2020-21 (M = 0.81, SD = 0.60 hours) compared to the number of hours per week in 2021-22 (M = 1.46, SD = 0.71 hours), a significant mean reduction of 0.65 hours, 95% CI [0.25, 0.47], p = <0.001.

The hypothesis that there will be a significant reduction in the hours of practical work completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic is partially upheld. The difference in the number of hours carried out pre-pandemic (M = 1.50, SD = 0.75 hours) compared to 2021-22 (M = 1.46, SD = 0.71 hours) was not significant (0.04 hours, 95% CI [-0.03, 0.06], p = 0.96.

Due to the kurtosis in the raw data, the repeated measures ANOVA was conducted with data from all three variables that had been transformed by applying a square root. This provided the same overall results which are included in Appendix A.

Respondents to the survey open items and those interviewed commented about less practical work taking place during the pandemic in 2020-21 and that there has been a return to pre-pandemic amounts during this academic year, 2021-22.

> *Typically, I would phone up a centre, and they would have done very few practicals and sometimes none at all. (Physics monitor)*
>
> *So, I certainly think, looking at the pandemic compared to pre-pandemic, schools have done a lot less practical work. (Chemistry monitor)*
>
> *Lots of centres are back on track doing as much practical work as before. (Biology monitor)*

The number of hours of practical work and school type, state or independent

Questions in the survey asked teacher respondents how many hours of practical work students completed per week before the pandemic, in 2020-21 and in 2021-22 and teachers were also asked to identify their type of school or college. A two-way mixed ANOVA was conducted to determine whether there was a significant interaction between the number of hours of practical work carried out per week at the different points in time compared with independent and state school and college.

There were outliers, as assessed by boxplot, and the data did show kurtosis at the 2021-22 time point for independent schools and at all 3 time periods for state schools, as assessed by calculation of the kurtosis. The two-way mixed ANOVA was conducted with and without the outliers. There was little material difference in the outcome as seen in Appendix B and so the analysis was carried out using the full data set.

There was homogeneity of variances ($p > 0.05$) and covariances ($p = 0.20$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 34.03, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.70$).

There was no significant interaction between the number of hours of practical work carried per week out at the different points in time compared with independent and state school, F(1, 60) = 0.05, p = 0.89, partial $\eta^2 = 0.015$. Figure 1 below shows this lack of interaction.

Due to the kurtosis in the data, the ANOVA was conducted on data from all three variables that had been transformed by applying a Log 10 as applying a square root was not sufficient to remove the effect of the kurtosis. This provided the same overall results which are included in Appendix B. The main analysis was conducted using the raw data for the mean to reflect the number of hours of practical work carried out per week.

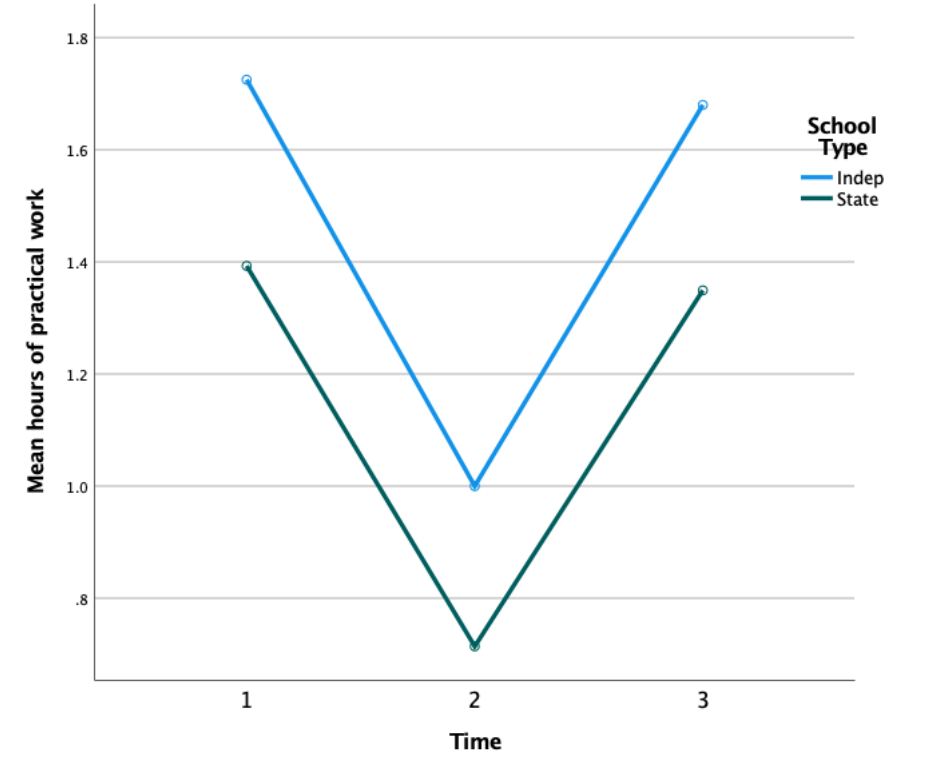


Figure 1. The mean hours of practical at each time point compared to school type

The main effect of time showed a significant difference in the number of hours of practical work carried out at the different time points, $F(1, 83) = 47.58$, $p < 0.001$, partial $\eta^2 = 0.44$. The main effect of school type showed that there was not a significant difference in the number of hours of practical carried out independent and state school $F(1, 60) = 3.93$, $p = 0.052$, partial $\eta^2 = 0.06$.

These data support the hypothesis that students from independent schools will have completed more hours of practical work per week before the pandemic, in 2021 and in 2022 in comparison to students from state funded schools and colleges.

The focus on the requirement for 12 practical activities

Teacher respondents and monitor respondents answered questions relating to the required practical activities from their perspective of assessing the CPAC directly or monitoring this assessment. In both cases, the responses were provided in the form of a Likert scale. Possible responses on this scale were strongly agree, agree, neither agree or disagree, disagree and strongly disagree. The higher the number (5) the greater the expression of disagreement and the lower the number (1), the greater the expression of agreement.

Table 3. Descriptive statistics of teacher focus on required practical activities pre-pandemic, in 2020-21 and in 2021-22

Descriptive statistics	Focus on required practical activities		
	Pre-pandemic	2020-21	2021-22
N	65	65	64
Mean	3.37	2.22	2.97
Std. Deviation	1.21	1.13	1.41
Range	4	4	4
Minimum	1	1	1
Maximum	5	5	5

Data in Table 3 shows that there was more of a teacher focus on required practical activities in 2020-21 than pre-pandemic and in 2021-22. Table 3 also shows that the focus on required practical activities in 2021-22 was similar to the focus before the pandemic.

Table 4. Descriptive statistics of monitors view on the focus on required practical activities pre-pandemic, in 2020-21 and in 2021-22

Descriptive statistics	Focus only on required practical activities		
	Pre-pandemic	2020-21	2021-22
N	27	27	27
Mean	3.00	1.70	2.30
Std. Deviation	1.07	0.61	0.91
Range	4	2	3
Minimum	1	1	1
Maximum	5	3	4

Data in Table 4 shows that monitors observed more of a focus on required practicals / specified practical work / PAGs / core practicals in 2020-21 than pre-pandemic and in

2021-22. Table 4 also shows that the observed focus on required practical activities was still more than the focus before the pandemic.

A one-way repeated measures ANOVA was conducted on both teacher and monitor data to determine whether there was a significant difference in the focus of practical work on required practical activities before the pandemic, in 2020-21 and in 2021-22.

There were no outliers as assessed by boxplot, as seen in Appendix D and no kurtosis as assessed by calculation of the kurtosis.

For the teacher data, the assumption of sphericity was met, as assessed by Mauchly's test of sphericity, χ^2 (2) = 0.276, p = 0.871. The pandemic resulted in a significant difference in the teacher focus on required practical activities between the three time periods investigated – before the pandemic, in 2020-21 and in 2021-2022 (F(2,126) = 40.79, p = <0.001).

Post hoc analysis with a Bonferroni adjustment revealed that there was an increased focus on required practical activities in 2020-21 (M = 2.22, SD = 1.129) compared to the pre-pandemic focus (M = 3.37, SD = 1.125), a significant mean increase in focus of 1.17, 95% CI [0.84, 1.50], p = <0.001. There was also an increased focus on required practical activities in 2020-21 (M = 2.22, SD = 1.129) compared to the focus in 2021-22 (M = 2.97, SD = 1.414), a significant mean increase in focus of 0.406, 95% CI [0.09, 0.72], p = 0.007.

For the monitor data, the assumption of sphericity was met, as assessed by Mauchly's test of sphericity, χ^2 (2) = 2.66, p = 0.265. The pandemic resulted in a significant difference in the observation of focus on required practical activities between the three time periods investigated – before the pandemic, in 2020-21 and in 2021-2022 (F(2,52) = 18.52, p = <0.001).

Post hoc analysis with a Bonferroni adjustment revealed that there was an increased observed focus on required practical activities in 2020-21 (M = 1.70, SD = 0.609) compared to the pre-pandemic observed focus (M = 3.00, SD = 1.074), a significant mean increase in focus of 1.30, 95% CI [0.77, 1.82], p = <0.001. There was also an increased observed focus on required practical activities in 2020-21 (M = 1.74, SD = 0.712) compared to the focus in 2021-22 (M = 2.30, SD = 0.912), a significant mean increase in observed focus of 0.60, 95% CI [0.11, 1.07], p = 0.012.

Responses to the survey open items and from those interviewed supported these data and comments were made about practical work during 2020-21 being mainly focused on required practical activities.

> *I think we've largely done well at A level to make the subject still be as practical as possible, but we have not done more than the 12 required practicals and have used some demonstrations. (Chemistry teacher, 11-19 maintained school)*
>
> *They didn't actually have to do all 12 but they needed to do sufficient for their students to be able to demonstrate competency routinely and consistently. But I know the schools that I spoke to wanted to get all 12 done in some way. (Physics monitor)*

The type of practical work completed by students

Several chi-squared tests were conducted to determine whether there was a significant difference in the types of practical work completed by students before the pandemic compared with students in the 2021 and 2022 cohorts. The results are in tables 5 and 6 below.

Table 5. Chi-squared test on teacher opinions about the types of practical completed by students

Type of practical work	χ^2	df	sig.
Method	0.61	2	0.735
Some investigation	8.59	2	0.014
Whole investigation	7.48	2	0.024
Teacher demonstration	0.23	2	0.891
Simulation	0.04	2	0.982
Fieldwork	11.39	2	0.003
Other	2.87	2	0.238

Table 6. Chi-squared test on monitor opinions about the types of practical completed by students

Type of practical work	χ^2	df	sig.
Method	2.10	2	0.349
Some investigation	11.37	2	0.003
Whole investigation	11.78	2	0.003
Teacher demonstration	7.56	2	0.023
Simulation	6.38	2	0.041
Fieldwork	4.48	2	0.101
Other	3.24	2	0.198

Some investigation work, whole investigations and fieldwork were the types of practical work that show a statistical significance in the opinion of teachers. This means the levels reported by teachers were significantly different before the pandemic, in 2020-21 and 2021-22. When looking at the raw data, shown in tables 7 to 9 below, it can be seen that there were fewer reports of all three of these types of practical activity completed in 2020-

21, according to teachers. Teachers reported that some investigation work and whole investigations were now, in this academic year, happening at similar levels to those seen before the pandemic but that fieldwork had not yet returned to the level seen before the pandemic.

Table 7. Data of teachers reporting that students were doing practical work that included some investigation

Time	Yes	No
Before the pandemic	62	8
2020-21	49	21
2021-22	59	11

Table 8. Data of teachers reporting that students were doing practical work that included a whole investigation

Time	Yes	No
Before the pandemic	24	46
2020-21	11	59
2021-22	23	47

Table 9. Data of teachers reporting that students were doing fieldwork

Time	Yes	No
Before the pandemic	25	45
2020-21	8	62
2021-22	17	53

Some investigation work, whole investigations, teacher demonstrations and simulations are the types of practical work that show a statistical significance in the opinion of monitors. This means the levels reported by monitors were significantly different before the pandemic, in 2020-21 and 2021-22. When looking at the raw data, shown in tables 10 to 13 below, there were fewer reports of all of these types of practical activity completed in 2020-21 other than teacher demonstrations which was higher in 2021-22, according to monitors. Monitors reported that some investigation work and simulations were now, in this academic year, happening at similar levels to those seen before the pandemic but that whole investigations had not yet returned to the level seen before the pandemic. Monitors

reported that teachers were now doing fewer demonstrations than they were in 2020-21 but not as few as they were doing before the pandemic.

Table 10. Data of monitors reporting that students were doing practical work that included some investigation

Time	Yes	No
Before the pandemic	25	2
2020-21	16	11
2021-22	24	3

Table 11. Data of monitors reporting that students were doing practical work that included a whole investigation

Time	Yes	No
Before the pandemic	14	13
2020-21	3	24
2021-22	6	21

Table 12. Data of monitors reporting that students were doing practical work that included teacher demonstrations

Time	Yes	No
Before the pandemic	10	17
2020-21	20	7
2021-22	14	13

Table 13. Data of monitors reporting that students were doing practical work that included simulations

Time	Yes	No
Before the pandemic	6	21
2020-21	1	26
2021-22	6	21

These data support the hypothesis that the type of practical work completed by students will vary for the 2021 and 2022 cohorts in comparison to before the pandemic. These data were further supported by responses to the survey open items and comments from those interviewed.

> *It was harder to find the time to complete a full investigation as we would do it in the summer of Y12 and school was mostly closed then. (Chemistry teacher, 11-19 academy)*

Some completed practicals as a class, others had to watch a teacher demonstration (often video-recorded and watched remotely). (Chemistry teacher, 11-19 maintained school)

The fairness of the assessment of practical skills in the A level sciences

Table 14 below shows the responses to the 26 items in the survey that asked for respondents' opinion about fairness on a Likert scale (strongly agree, agree, neither agree or disagree, disagree, strongly disagree). Each item was based on a sense of fairness as defined by Shaw & Nisbet (2020). The items were presented to respondents in a random order so that they did not see items about the same sense of fairness together.

Table 14. Data of Likert responses to fairness items

Fairness sense	Item – To what extent do you agree or disagree that during the pandemic fairness may have been reduced because…	Mean	SD	Strongly agree and agree (%)	Neither agree or disagree (%)	Strongly disagree and disagree (%)
Formal	The rules in place for the assessment of the CPAC were not always appropriately followed	2.66	0.97	50.6	29.6	18.5
	The rules in place for the assessment of the CPAC were sometimes broken	3.04	1.02	38.3	30.0	35.0
	It was not always possible to follow the rules in place for the assessment of the CPAC	3.16	1.11	32.1	24.7	42.2
	It was not possible to distinguish between students with weak or strong practical skills as fewer than 12 practical activities could be used to assess the CPAC	3.06	1.10	35.8	24.7	37.1
Implied contractual	The assessment of the CPAC was not carried out as students and teachers had legitimately expected	2.90	0.96	37.0	37.0	24.7
	Teachers legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic	3.11	0.96	28.4	34.6	35.8
	Students legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic	3.15	1.01	28.4	29.6	39.5
Relational	Not all students were treated the same way in the assessment of the CPAC	2.58	0.91	53.1	29.6	16.1
	Groups of students were treated differently to each other in the assessment of the CPAC	3.00	1.04	39.5	22.2	37.1
	Even prior to the pandemic, students have never all consistently had the same access to technician support when being assessed on the CPAC	2.83	1.06	40.7	27.2	28.4
	Even prior to the pandemic, students have never all consistently had the same access to specialist teachers when being assessed on the CPAC	2.90	1.00	35.8	33.3	29.6
	Even prior to the pandemic, students have never all consistently had the same access to teachers that are confident and experienced when being assessed on the CPAC	2.80	1.05	42.0	28.4	28.4
	Even prior to the pandemic, students have never had the same access to practical resources when being assessed on the CPAC	2.74	1.12	36.9	24.7	27.2
	Students could have been assessed on a number of practical skills that varied widely	2.96	0.92	35.8	34.6	28.4

	Students could have been assessed on types of practical skills that varied widely	2.85	1.03	44.4	25.9	28.4
	This cohort of students were not treated in the same way as previous cohorts of students in the assessment of the CPAC	2.66	0.95	51.8	28.4	18.5
Retributive	It was not always possible to appropriately reward students for their competence in their practical skills	2.94	1.00	37.0	29.6	30.8
	The result of the endorsement did not always reflect student competence in practical skills	2.84	1.07	44.5	25.9	27.2
	The assessment of the CPAC did not always reward students for their competence in practical skills	3.29	0.96	23.4	30.9	44.4
Consequential	Students will be treated unjustly in the future based on their result in the Practical Endorsement	3.55	0.84	11.1	33.3	44.3
	Students will be treated unjustly in the future based on their practical skills as assessed by the CPAC	3.48	0.90	16.0	27.2	55.5
	Students with weak practical skills could use their pass in the Practical Endorsement to access university courses where this is a requirement	2.89	0.98	35.8	37.0	25.9
	Students with weak practical skills will have opportunities that they would not have had if their practical skills had been assessed as required prior to the pandemic	2.76	0.87	39.5	38.3	19.7
	Students with strong practical skills will not have the opportunities they would have had if all students had been assessed as required prior to the pandemic	2.90	1.07	44.4	22.2	32.1
Retrospective	The assessment of the CPAC has never been a fair assessment method of practical skills	3.38	1.13	25.9	23.5	49.4
	The award of a pass in the Practical Endorsement has never been a fair reflection of competence in the skills assessed by the CPAC	3.29	1.16	28.4	25.9	44.5

Table 14, in descriptive terms, provides clear information about the views of fairness provided by the survey respondents. For example, more respondents agreed (50.6%) than disagreed (18.5%) that the rules in place for the assessment of the CPAC were not always appropriately followed. Also, more respondents disagreed (49.4%) than agreed (25.9%) that the assessment of the CPAC has never been a fair assessment method of practical skills.

Scales were created of these data to provide more reliable measures. A Cronbach's Alpha was calculated to measure the internal consistency of these scales to provide information as to whether the items are measuring a common construct. The results are in Table 15 below.

Table 15. Cronbach's Alpha test on fairness items

Sense of fairness	N	Items	Cronbach's α
Overall fairness	81	26	.941
Formal fairness	81	4	.768
Implied contractual fairness	81	3	.793
Relational fairness	81	9	.843
Retributive fairness	81	3	.584
Consequential fairness	81	5	.776
Retrospective fairness	81	2	.809

Included in this test was analysis to assess the value of removing items to provide a better result from this test. It was determined that these values show acceptable internal consistency as all, apart from retributive fairness, give a value above 0.7 (Field, 2018 p. 823-4). The retributive fairness value of 0.584 is a less acceptable internal consistency value. There may have been a reason such as the low number of items (3) that may have affected the internal reliability (Field, 2018 p.823).

As these values show acceptable internal consistency in all but one sense of fairness, it was appropriate to create scales of each sense of fairness to use in further statistical analysis. The scales were created by adding the values of items within each sense of fairness and dividing by the number of items assigned to that sense of fairness. This was done in order to return the values to the same scale as used in the raw data – 1 (strongly agree) to 5 (strongly disagree).

When assessing the data within each scale for the assumptions to conduct a one way repeated measures ANOVA test, the data showed severe kurtosis even when the raw data was transformed to the square root and Log10 values as seen in the tables in Appendix D. It was determined therefore that a Friedman test was an appropriate non-parametric test to establish if there were any significant differences between the distributions of the fairness items.

A Friedman test was conducted to determine if there were significant differences in views to each sense of fairness. The differences were statistically significant, $\chi^2(5) = 37.009$, $p = <0.001$. Respondents to these survey items about fairness identified differences between some of the senses of fairness identified in Nisbet and Shaw's typology.

Post-hoc tests were conducted. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Statistical significance was accepted at the $p < 0.003$ level. Respondents to the survey identified a significant difference between relational and consequential fairness ($p < 0.001$), relational and retrospective fairness ($p < 0.001$) and formal and retrospective fairness ($p < 0.001$).

Respondents to the survey open items and interviewees mentioned relational fairness, the fair treatment of students with those in their cohort, in a different school or in a different time period, most frequently.

> *Teachers generally did their best to ensure the CPAC assessment ran as normal but some centres/students had a different experience due to the pandemic. (Physics monitor)*
>
> *Centre resources and experience varied enormously. (Biology monitor)*
>
> *Absence has made it difficult to assess all students against the same skills. (Biology teacher, 11-19 academy)*

However, this was often followed up with comments that the circumstances of the pandemic were no-one's fault and therefore it was not appropriate to consider it unfair. Teachers, students, and exam boards had done the best they could in the circumstances of the pandemic to make the provision of practical work and its assessment as fair as possible for A level science students.

> *I don't think there's 'unfairness' - everyone worked with what they had. (Physics teacher, 11-19 academy)*
>
> *As fair as possible in the circumstances. Teachers are trusted to use professional judgement. (Chemistry teacher, further education college)*
>
> *Students were assessed on the practical work that had been carried out even though many had minimal opportunity to develop practical skills, it was fair not to penalise them. (Biology teacher, 11-19 academy)*

Fairness and the hours of practical work per week

Questions in the survey asked teacher respondents how many hours of practical work students completed per week before the pandemic, in 2020-21 and in 2021-22. A Pearson's correlation was conducted to determine any relationship between this

information and respondents' views of the different senses of fairness. The results of this correlation are in Table 16 below.

Table 16. Pearson's correlations between fairness scales and the number of hours of practical work students completed per week before the pandemic, in 2020-21 and in 2021-22

	Correlation co-efficient			sig.		
	Pre-pandemic	2021	2022	Pre-pandemic	2021	2022
Overall fairness	-0.23	-0.19	-0.14	0.07	0.14	0.27
Formal fairness	-0.10	-0.05	-0.10	0.45	0.68	0.44
Implied contractual fairness	-0.23	-0.09	-0.15	0.07	0.49	0.24
Relational fairness	-0.28	-0.26	-0.30	0.03	0.04	0.02
Retributive fairness	-0.12	-0.15	-0.14	0.36	0.25	0.27
Consequential fairness	-0.30	-0.21	-0.15	0.02	0.10	0.19
Retrospective fairness	-0.19	-0.04	-0.19	0.14	0.73	0.13

The data in table 16 show that all correlations were negative: the fewer hours of practical work completed by students per week, then the more each sense of fairness was impacted. However, all correlations were weak.

The correlation between pre-pandemic number of hours of practical work and consequential fairness was significant ($r(64) = -0.30$, $p = 0.02$). The fewer hours that were spent on practical activities before the pandemic, the more likely the respondents thought that consequential fairness, the possibility that the outcome of the endorsement would be used for unfair actions in the future, would have an impact. For example, the fewer hours of practical work, the more likely a student would have weaker practical skills than indicated by their pass in the endorsement which might lead them to be able to access a higher education course or employment on the basis of their pass. Their practical skills

may not be as good as would be needed either for undergraduate study or to carry out the requirements of a job.

> *Some students, with perhaps less good practical skills will still be accessing higher education, whereas in the past, or perhaps in the future, they'd have to be better to access higher education. (Chemistry monitor)*

The correlations between the number of hours of practical work at all time periods and relational fairness were significant (pre-pandemic ($r(64) = -0.28$, $p = 0.03$), in 2020-21 ($r(64) = -0.26$, $p = 0.04$) and in 2021-22 ($r(62) = -0.30$, $p = 0.02$)). Teachers and monitors therefore thought that the fewer hours that were spent on practical work at all points in time, pandemic or not, relational fairness was impacted. If students completed a different number of hours of practical work each week, they were being treated differently. Relational fairness was frequently highlighted in response to the survey open items and by interviewees. Interestingly, comments were often followed up with comments that no-one was to blame for the pandemic so any different treatment of students was not really an issue of fairness.

The data about retributive fairness was not significant. Teachers and monitors believed that a pass in the endorsement was deserved even though the pandemic had a significant impact on the number of hours of practical work and on the type of activities completed by students. Comments in response to the survey open items and by interviewees supported these data.

> *I don't think the CPAC distinguish between students of different abilities, rather it is an acknowledgement of a basic level of competence. They either pass or it is not reported. Because so much practical work is done in comparison with pre-2015, students are more likely to have that basic competence. Perhaps the level of distinction is at a very low level - most students pass. I don't think the pandemic made a great difference to this. (Physics monitor)*

In addition, teachers and monitors did not think that formal fairness would have been impacted to any significant level even though the rules for the assessment of the CPAC were relaxed during the pandemic. It had not been hard to judge what competence in the CPAC really meant from one student or school and college to another. Comments in response to the survey open items and by interviewees support these data

> *"I think that given the circumstances, IF the rules were followed then it was as fair as it could have been." Chemistry teacher, independent school*
>
> *"I do not feel that the staff took advantage of the reduced number of required practicals to unfairly benefit their students." Physics monitor*

Measuring Nisbet and Shaw's typology of fairness

An exploratory factor analysis using principal components extraction and an oblimin rotation was conducted because it was expected that the factors would correlate. The table of data is in Appendix E.

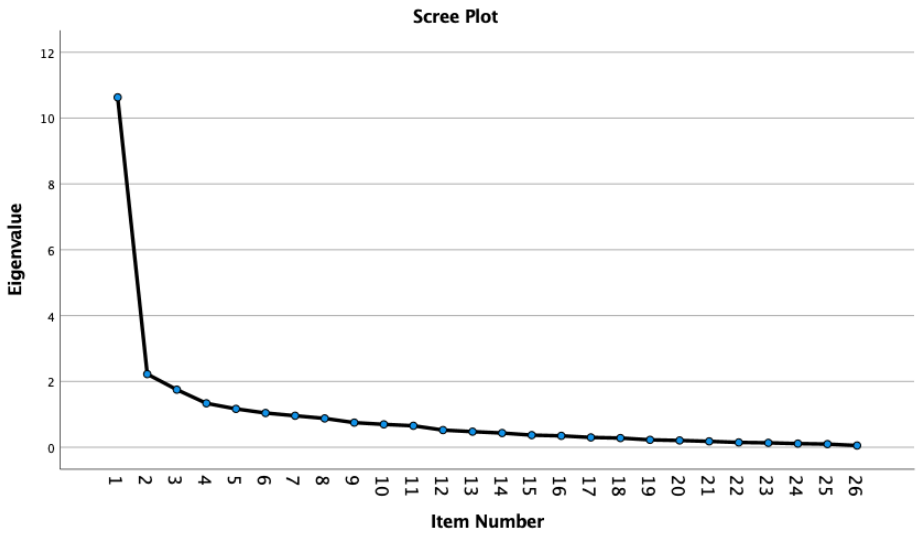


Figure 2. Scree plot of eigenvalues of fairness items

There were six factors with eigenvalues over one although the scree plot (figure 2 above) suggested just one or two factors, with additional factors not adding much explained variance. All items loaded onto the first factor (Appendix E). The lowest loading was -0.50. Items relating to relational fairness tended to load onto factor two but with loadings which were lower than those for factor one. This suggested that a one factor model provided the best fit to the data. This factor analysis, therefore, did not support the distinctions between senses of fairness suggested by Nisbet and Shaw's typology.

Table 1. Coding of survey open items and interview responses

Node	Frequency of references
Overall fairness	16
Formal fairness	28
Implied contractual fairness	6
Relational fairness	78
Retributive fairness	21
Consequential fairness	2
Retrospective fairness	11
As fair as it could be in the circumstances	20
Impacts on practical work	30
Amount of practical work	33

Table 1 above shows that responses to the survey open items and by interviewees showed that most respondents were concerned about relational fairness. Comments were based on whether the pandemic caused the assessment of practical skills to be unfair when comparing the experience of one student with another student. Most respondents commented that the impact of the pandemic would have resulted in a large variation in the experiences that students may have had in relation to provision of practical work.

Unfair due to vast differences in resources, staffing, experiences and expectations between cohorts and centres. (Biology teacher, independent school)

Other physics teachers teaching A level had little access to practical space. So, that was the first thing which might impact fairness in relation to how similar provision was for different people in different places. (Physics teacher, faith school)

Some respondents did say that it would still have been fair if comparing students.

Fair because students just need to demonstrate a certain level of competence. (Chemistry teacher, 11-19 academy)

Discussion

Impact of the pandemic on the delivery of practical work

To determine the impact of the pandemic on the assessment of A level science practical skills and any impact on the fairness of the award of the Practical Endorsement, it was important to establish any effect of the pandemic on teachers' ability to deliver practical work to their students.

Teachers and monitors reported a significant reduction in hours of practical work delivered to students in 2020-21, the academic year that had the most disruption to education in the form of partial school closures and remote learning. In 2020-21, remote learning due to the national lockdown had an inevitable impact on the delivery of practical work. The science curriculum, with its significant practical element, had to be adapted, and some schools were not able to include the practical part in delivery (Ofsted, 2020a). Adaptations included teachers demonstrating practical work in science rather than students being able to carry it out themselves (Ofsted, 2020b).

Once the lockdown was over and schools and colleges were fully open, public health restrictions prevented full access to practical work for many students. These restrictions included practical science equipment being quarantined (CLEAPSS, 2020), science technicians being redeployed to other pandemic related roles such as student testing for COVID-19 and the approach to 'bubbles' (restricting movement of students around schools and keeping them with their cohort as far as possible (Southall et al, 2021)). The 'bubbles' meant that students may not have had access to laboratory space (Ofqual,

2021c). Throughout this academic year, staff and student absence due to pandemic related illness continued to impact the delivery of the full curriculum, including practical science lessons (Southall et al, 2021). This included substantial variation in attendance by teachers and students across regions of the country because of COVID-19 (Shaw & Nisbet, 2021). Some teachers reported the additional expectation that students at home due to COVID-19 infection would dial into lessons remotely. Teachers' attention was then spread across students in the room and at home so delivering a practical lesson in these circumstances was simply too challenging.

Teachers had to focus on ensuring that the theory part of courses was covered in order for students to be prepared for assessments to provide evidence for the Teacher Assessed Grades (TAGs) that were used to provide A level grades in summer 2021 (Ofqual, 2021e). Teachers and monitors reported that the collection of this evidence resulted in practical work being less of a priority for face to face time with students. This approach to grades resulted in students being assessed on what they had been taught as teachers could tailor the assessments used as evidence. This may have resulted in fewer questions that indirectly assess practical skills (Abrahams et al, 2013) being used in the assessments as students may have not had the opportunity to complete all the required practical activities. Comments to the survey open response items and by interviewees suggested that this was a fair approach.

Teachers and monitors did not report a significant reduction in the number of hours of practical work in 2021-22 in comparison to pre-pandemic. When responding to the survey open items and in the interviews, they explained that in spite of continued teacher and student absences due to COVID-19, there had been an effort to return the amount of

practical work delivered to A level students to that delivered before the pandemic. This is a very encouraging finding especially when considering the overwhelming opposition to the endorsement as an approach to the assessment of A level practical science skills by the science community. Teachers really do value it as an essential part of the course and were keen to ensure students had as many opportunities as possible to complete practical activities. It is also likely that there was a shift in focus with the return of summer exams and the need for students to be prepared to answer the exam questions based on practical skills. Teachers reported that it was important that students had the hands-on experience to be able to tackle these questions with confidence.

Reports from teachers and monitors suggested that independent schools conducted more practical work than state schools and that this was the case across each of the three time periods studied. Reasons for this included that the disruption caused by COVID-19 infection within bubbles (this caused bubbles e.g., a whole year group, to be sent home from school or college) appeared to affect state schools more than independent schools (Ofqual, 2021b). This would have impacted on the delivery of practical work as it may have meant that science teachers were absent as well as students having to learn remotely. Those teachers would then not have been able to deliver practical lessons to students even if those students were still able to attend school or college. In addition, the same report documents that independent school students had access to more hours of remote learning in all periods when schools and colleges were closed to most students. It is reasonable to suggest that this would have resulted in more time to catch up on A level science practical work when lockdowns were over as there would have been opportunity to cover more theory content during remote learning.

The A level science outcomes by centre type, including state and independent, are shown below in figures 3 (biology), 4 (chemistry) and 5 (physics). It is not possible to see this data for A level geology as Ofqual publish data for this subject within a category called 'other sciences' as the entries are small in comparison to the biology, chemistry and physics entries.

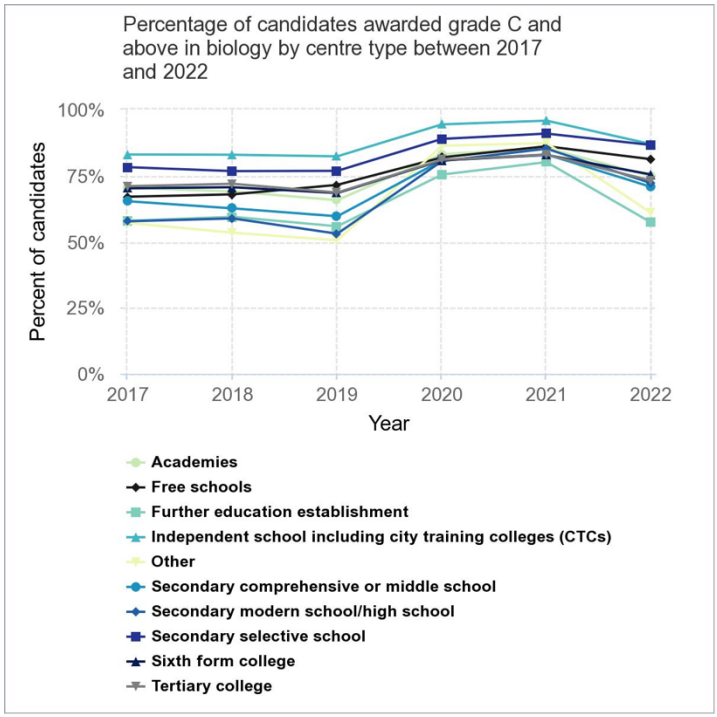


Figure 3. Percentage of students awarded grade C and above in A level biology by centre type between 2017 and 2022 (Ofqual, 2022)

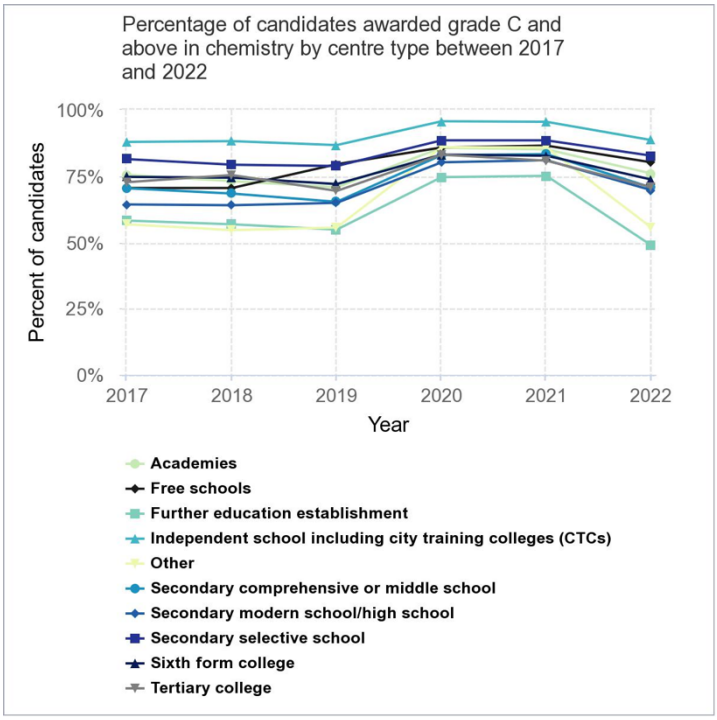


Figure 4. Percentage of students awarded grade C and above in A level chemistry by centre type between 2017 and 2022 (Ofqual, 2022)

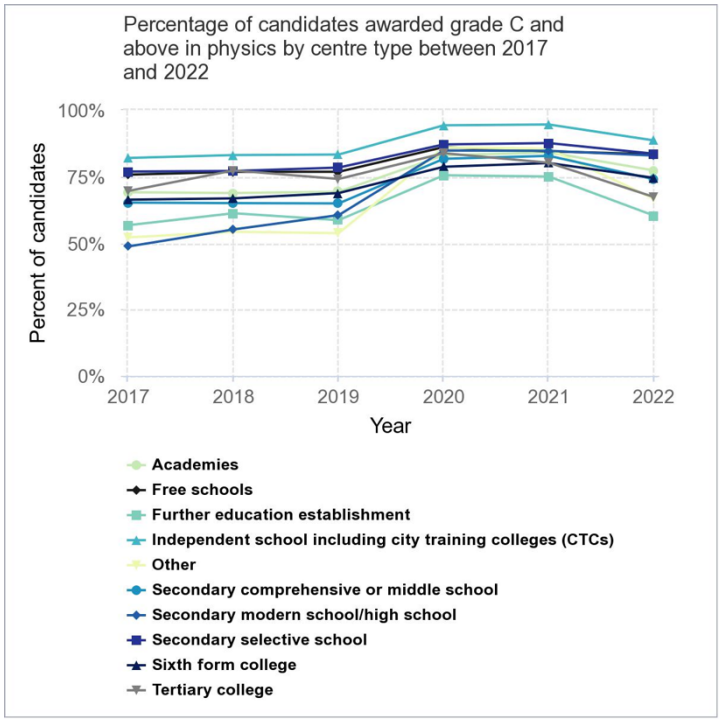


Figure 5. Percentage of students awarded grade C and above in A level physics by centre type between 2017 and 2022 (Ofqual, 2022)

In all three A level science subjects above, the data shows that more students from independent schools achieve a grade C or above than in all types of state school. The analysis in this project shows that independent school students have had access to more hours of practical work before the pandemic, in 2020-21 and 2021-22 and this may have contributed to higher overall grades. This access may have resulted in students from independent schools being more successful when answering the exam questions that indirectly assess practical skills (Abrahams et al, 2013). However, it is important to note that many factors contribute to the grades achieved by A level students.

Teachers and monitors reported that there was more of a focus on the 12 required practical activities for the assessment of the CPAC in 2020-21 than before the pandemic and in 2021-22. This was expected as the rules were relaxed so that it was not necessary to do all 12 activities if teachers could assess the CPAC on fewer activities (Ofqual, 2020a). This relaxation of the rules remained in place for the assessment that took place for students entering for A level qualifications in summer 2022 (Ofqual, 2021a) so it was expected that there was still more of a focus in this last time period than before the pandemic.

The reasons for the focus on the 12 practical activities were likely to be similar to those described above for reasons for fewer hours of practical work in 2020-21 such as a lack of time and resources to focus on practical work. In addition, as exams were planned for summer 2022, it was likely that teachers were focused on ensuring that all of the theory content had been taught.

Teachers and monitors reported that there was less of a focus on investigative practical work, either some investigation or a whole investigation in 2020-21 than before the pandemic and that whole investigations were still, in 2021-22, not as frequent as before the pandemic. The reasons were likely to be similar to those described above for reasons for fewer hours of practical work in 2020-21 such as a lack of time and resources to focus on practical work. It might have been expected that there would still have been a focus on some investigative work as this is one of the CPAC and so students need to be able to demonstrate competency in this area to be awarded the Practical Endorsement. Teachers reported that this part of the CPAC is time and resource intensive so it was understandable that it proved challenging during the pandemic.

Impact of the pandemic on the fairness of the assessment of practical skills

Two interpretations of the analysis on the senses of fairness in Nisbet and Shaw's typology (2020) were possible from the data. The first interpretation was that respondents to the survey did not clearly identify the six different senses of fairness. In the context of the pandemic, fairness was not nuanced fully to these different senses.

To consider whether respondents did distinguish fully between the senses of fairness, an exploratory factor analysis was conducted. A one factor model provided the best fit for the data and so, in the context of the assessment of practical skills, did not support fully the distinctions between the senses of fairness as suggested by Nisbet and Shaw's typology (2020). As survey respondents did not clearly separate all of the senses of fairness in the typology, it may mean that the analysis of the typology in the context of practical skills and the impact of the pandemic was weak. It is important to emphasise

that the context of this analysis as another set of circumstances may result in distinctions between all the senses of fairness. This possibility cannot be ruled out.

The typology can be seen as an academic analysis of fairness in assessment and so there may have been a lack of comprehension by survey respondents. Comments did indicate that teachers and monitors were not always sure that their views were related to fairness and concern was expressed that respondents were not clear for whom the assessment of practical skills might have been unfair. In addition, there may have been item fatigue with the survey items asking about the different senses of fairness as comments did suggest that respondents found some items repetitive or confusing.

It is worth noting that respondents to the survey did identify relational fairness when answering the open questions. Comments emphasised an awareness of much variation in the way that students were treated with respect to the assessment of practical skills, both between schools and colleges in the same year and in different years due to the impact of the pandemic. This may have been due to different responses to the pandemic overall by individual schools and colleges which in turn impacted on the ability to deliver practical work.

The second interpretation was that the senses of fairness, where identified, were not differentially impacted over the course of the pandemic. It might be the case that if another set of circumstances caused changes to the assessment of practical skills, the different senses of fairness might actually become more important.

Respondents to the survey said that relational fairness was impacted more than consequential fairness, that relational fairness was impacted more than retrospective fairness and that formal fairness was impacted more than retrospective fairness as shown by the Friedman test. Comments in response to the open survey items and by interviewees focused on relational fairness and the fair or unfair treatment of students if their opportunities for practical work were very different due to the pandemic. There was also an awareness from respondents of the impact of following the rules for the assessment of the CPAC, formal fairness. The comments about retrospective fairness, to do with actions of the past, by respondents were focused on whether the endorsement was ever a fair method of assessment of practical skills and this tended not to be linked to the impact of the pandemic.

The negative correlations between the fairness scales and the number of hours of practical work at all time periods showed that respondents thought that the fewer hours of practical work, the more all senses of fairness were impacted. This was the case whatever the circumstances, including a pandemic. There was little distinction between the fairness scales in the analyses and the differences were generally not significant. This suggests that respondents did not, in most cases, consider fairness to be differentially impacted which may have been for the reasons described above.

Consequential fairness was one exception and this did correlate significantly with number of hours of practical work carried out before the pandemic. This suggested that consequential fairness had more of an impact than the other senses of fairness. The fewer hours of practical work, the more likely a student would have weaker practical skills than indicated by their pass in the endorsement which might lead them to be able to access a

higher education course or employment on the basis of their pass. Their practical skills may not be as good as would be needed either for undergraduate study or to carry out the requirements of a job. The experience of the endorsement over the last seven years suggested that teachers and monitors viewed a pass as communicating a standard of competency in students' practical skills that needed to be maintained. If less opportunity for practical work leads to weaker skills then this would be considered to be consequentially unfair.

The correlation between consequential fairness and hours of practical work was not significant in 2020-21 or 2021-22 which suggested that respondents did not perceive this to be unfair in the circumstances of the pandemic. If students had completed less practical work and had weaker skills, it was because of the impact of the pandemic which was not anyone's fault and, in fact, teachers had done their best in the circumstances. Students with weaker skills might have accessed related higher education courses or employment where a particular standard of practical skills was expected but this was not unfair in these circumstances. Of course, there was a presumption that higher education and employers would be providing support for students in these circumstances. Although not mentioned in any comments, it would only be fair if higher education and employers do step up and provide support due to the context of the pandemic and its impact.

Relational fairness and hours of practical work were significantly correlated at all time periods. This suggested that respondents were always of the opinion that fewer hours of practical work at any point in time would be judged to be unfair comparing students with each other, either between schools and colleges or when the endorsement is awarded in different years. Again, this outcome for relational fairness may support the greater impact

of this sense of fairness as respondents were clear that there were fairness issues when considering whether students were treated the same and had the same practical opportunities.

The correlations for formal fairness at all time points were not significant. This showed that teachers and monitors did not identify this as having a greater impact than the other senses of fairness. This lack of significance was somewhat surprising as 50.6% of respondents agreed or strongly agreed that the rules for the assessment of the CPAC had not always been appropriately followed. However, the rules had been relaxed so that it was not necessary to do all 12 practical activities for 2020-21 (Ofqual, 2020a) and for 2021-22 (Ofqual, 2021a). This suggested that respondents may have thought the rules were not always followed but this did not necessarily mean that the assessment was unfair. The rules had been changed to address the issue of fewer opportunities for practical work, an impact of the pandemic, which was seen as a greater unfairness. It may have been unfair if students were not awarded the endorsement due to not being able to do sufficient practical work according to the pre-pandemic rules through no fault of their own.

It is worth noting that there were quite a large number of correlation calculations and every time one was conducted, there was a 5% chance of a significant effect that was actually down to chance. Any significance could just have been noise in the data and a consequence of doing multiple correlations. Therefore, it was important to consider the significance for consequential fairness and relational fairness (which might have supported the notion of different constructs in the scales) but this significance could have been a product of multiple testing.

Overall, if the scales for the different senses of fairness had performed quite differently at the different points in time, it would imply that they are measuring different constructs but as they performed similarly, it casts further doubt on whether the respondents to the survey did distinguish between the impact of the senses. It is interesting to have an academic debate about e.g., formal vs relational but in the context of the fairness of the assessment of practical skills and the pandemic, the respondents did not identify many differences in their impact.

Fairness in the circumstances of the pandemic

The most prevailing opinion expressed in the survey open items and by interviewees was that in the context of the assessment of practical skills and the pandemic, fairness was binary – the assessment was either fair or it was not fair and teachers and monitors did not see the need to explore any nuance or define the senses of fairness because, and this was a frequently made point, blame for the pandemic could not be attributed anywhere. The assessment was as fair as it could be in the circumstances and that was all that needed to be said. No-one intended that it be unfair.

A similar view was expressed about reliability by He, Boyle & Opposs (2011) when reporting "The findings suggested a demarcation in the minds of the public between inevitable errors in the assessment process and preventable errors. The research participants appeared to accept that a certain amount of error was inevitable in a large examinations system, but they could be intolerant of 'preventable errors'." This opinion was also expressed by Chamberlain (2013) when discussing interviews with a range of

stakeholders about error relating to reliability in assessment. If error was caused by a particular approach to the sampling of questions, the stakeholders did not view this as unfair as it was no-one's fault but they were intolerant of perceived human error where responsibility could be attributed to a person.

This is exactly how the respondents to the survey open items and interviewees viewed fairness – the assessment of the CPAC may have been unfair but it was acceptable as it was not preventable and everyone had done their best in the circumstances.

The application of Nisbet and Shaw's typology to this example of the fairness of the assessment of practical skills in the A level sciences as a result of the pandemic was partially successful. Respondents to the survey and interviewees did identify some differential impacts of the senses of fairness described and did have a view about, for example, whether students would be able to access higher education with weaker practical skills than the endorsement would suggest (consequential fairness). Also, that students might have expected their learning in an A level science subject to be more based around practical work that was then not possible due to the pandemic (implied contractual fairness).

Respondents and interviewees did not identify any other senses of fairness that could be separated from those in the typology so in the context of this research, the typology is exhaustive. Nonetheless, in the view of the participants the circumstances of the pandemic were such that any unfairness was no-one's fault and under no-one's control so it was not really acceptable to class it as unfair.

Fairness, validity and reliability

The introduction to this project considered the relationship between fairness, validity and reliability. The most recent edition of the *Standards* describes these three concepts as the 'big three' of educational assessment and comments that "Fairness is a fundamental validity issue…" (AERA et al, 2014). Nisbet & Shaw (2020) also discuss the relationship and conclude that fairness is an essential requirement for validity. They add that there is more to validity than fairness and that it is possible for a fair test to not be valid." (Nisbet & Shaw, 2022). Shaw & Nisbet (2021) linked fairness closely with reliability when they said, "Fairness is closely tied to the concept of reliability and any threat to reliability will call the fairness of the test into question."

This research has highlighted the concepts of validity and reliability as well as fairness. For example, if students did not complete all 12 practical activities as part of their A level science course and were still awarded the Practical Endorsement, this may have impacted validity in terms of construct underrepresentation. The *Standards* defines this as "the extent to which a test fails to capture important aspects of the construct domain that the test is intended to measure, resulting in test scores that do not fully represent the content." (AERA et al, 2014). In the context of the assessment of practical skills, it might be significant which activities were not completed by students. For example, an A level chemistry teacher might be considered to have allowed construct underrepresentation if a student was awarded a pass in the endorsement without having completed a titration. This may call the validity of that pass into question.

The results above indicate that the pandemic caused a differential opportunity to access practical work due to school closures, absence and public health restrictions. This had a relational fairness application – students were not treated the same way with respect to access to practical work as there was a wide variation in opportunity to complete practical work, in some cases at all and in others to the extent to which the CPAC could be assessed fairly. In the perception of respondents to the survey, the change to the rules for the assessment of the CPAC was a fair response to this application.

This line of thinking can be extended to consider whether the assessment of A level practical science skills is ever fair based on the different approaches and experience of teachers and the prominence of practical work in teaching and learning. Some teachers prefer to teach through the medium of practical work and others do not. Some have a lot of confidence and experience, others do not.

The pandemic has put a spotlight on this sort of variation, often ignored when discussing high-stakes assessments – there are differences in the teaching and learning that groups of students experience which results in a discussion about fairness in the context of an opportunity to learn (Shaw & Nisbet, 2021). There has also been more emphasis on the variation in outcomes of students from different socio-economic backgrounds and between the north and south regions of England (Sutton Trust, 2022).

There is no doubt, when considering the responses to the survey open items and comments by interviewees, that teachers and monitors viewed the endorsement to be as fair as it could be in 2021 and 2022 when less evidence was needed and varying amounts of evidence were collected. This is retributive fairness. “A fair assessment outcome is

seen as “deserved” because of the hard work that the student has done or because it matches some other evidence of the student’s ability.” (Shaw & Nisbet, 2021). Even though there was less or varying evidence, it was still fair as it reflected the students’ skills. When looking at the fairness items, 44.4% of respondents disagreed or strongly disagreed with the statement that “the assessment of the CPAC did not always reward students for their competence in practical skills” indicating that more respondents were of the view that students were rewarded.

The context of the pandemic was the reason teachers and monitors did not want to label the assessment of practical skills as unfair. Nisbet & Shaw (2022) extend the idea above that a fair assessment is ‘deserved’ when seen in context. There was widespread adaptation of assessment arrangements for GCSE and A level qualifications in 2021 and 2022, including for the assessment of practical skills. These may still be considered to be fair, even if there was a lack of equivalence for students with different experiences. The adaptations were designed to allow students the opportunity to show what they knew and could do (Nisbet & Shaw, 2022). Relational fairness was not the primary concern for respondents and interviewees as that could have led to students being penalised. The most important aspect was that students were rewarded for what they had been able to show and not disadvantaged due to circumstances that were outside of their control and no-one’s fault.

The application of fairness to policy and practice

The views expressed in this research suggested that the most important consideration was that students were not disadvantaged by the impacts of the pandemic. Efforts were made by policy makers to ensure that the rules allowed teachers to award a pass in the Practical Endorsement in the circumstances, even if this meant less or varying amounts of evidence than normal. When qualifications are reformed, Ofqual uses comparable outcomes to ensure that students sitting assessments in the first years of the new qualifications are not unfairly disadvantaged (Ofqual, 2017). This disadvantage in the first years of new qualifications might be the result of a lack of familiarity with the assessments, fewer resources and support, fewer past exam papers, a lack of teacher confidence. Comparable outcomes uses statistical evidence and the opinions of senior examiners to ensure that outcomes do not unfairly disadvantage students, assuming that the national cohort has a similar prior achievement profile to that in previous years. Interventions for practical science assessment during the pandemic protected the pass results of students even when their skills may not have been as strong.

In any year, students are offered support in the form of reasonable adjustments and special consideration (JCQ, 2022). The Equality Act 2010 means that students with disabilities that might affect their access to assessments are allowed reasonable adjustments to allow fair access which are put in place in advance of the assessments. Special considerations are applied after assessments have taken place if students ability to perform to the best of their ability has been affected by illness, bereavement or other similar circumstances. This support is well understood by the public including schools and colleges, teachers, students and their parents. The changes to the assessment of A

level practical science skills fall into a similar category as reasonable adjustments as the changes were intended to allow fair access to an endorsement result whatever the circumstances of the pandemic. The public accepted these as fair.

The outcome of this research is an example of the challenge there is for Ofqual and others working in the qualification system to know what might be considered fair to users of qualifications and in the perception of the public. The change to the rules of the assessment of the CPAC was acceptable and not considered to be an issue of fairness and in fact, respondents would have found it less acceptable for the rules to have remained the same based on the difficulties with the delivery of practical work. In the context of the pandemic more widely, the approach to the standardisation of centre assessed grades in the summer of 2020 was not viewed to be acceptable by the public (Ofqual, 2020b). Those involved in the qualification system must consider the work that needs to be done to consider public beliefs and perceptions about fairness in order to be more aware of responses to changes that might be needed in the future.

Conclusion

The COVID-19 pandemic affected the amount of practical work delivered to A level science students. School closures, absences and public health restrictions all contributed to fewer practical opportunities, particularly in the academic year 2020-21. There has been quicker and more comprehensive recovery in this academic year, 2021-22 than expected. Independent schools provide more practical work for students than state schools.

All senses of fairness as described by Nisbet & Shaw (2021) were identified by respondents to the survey and interviewees in relation to the assessment of practical skills. However, teachers and monitors did not see any merit in discussing them at great length as the circumstances of the pandemic had resulted in the assessment being as fair as anyone could have expected. It was acknowledged that this was different to normal times, when some students do have a poorer experience, as the pandemic impacted everyone and had national implications, although not consistent effects in all parts of the country, and so it was appropriate to change the rules and make adaptations that offered support to all students.

Limitations

Limitations of this project included a difficulty in accessing A level science teachers to ask them to complete the survey so the sample size is not as big as the author would have preferred. As respondents to the survey were usually of the opinion that the assessment of

practical skills had been as fair as possible, there was a degree of fatigue over the number of items asking about fairness.

Respondents also may have lacked an understanding of the senses of fairness. Information about these senses could have been included in the survey but this might have biased the responses and the reading would have increased the cognitive load. There seemed to be a lack of clarity about fair to whom – "It is often assumed that fairness applies only to candidates. But there are others whose interests may also be at stake, for example, candidates' peers (who did not take the test), users of the assessment outcomes (e.g., employers or universities) or society at large." (Shaw & Nisbet, 2021). As a result, respondents were unsure in their comments.

Further research

To understand the long term consequences of the changes to the rules, further research could survey teaching fellows of first year undergraduate students about the practical skills as a result of the pandemic. This could be an extension to the work carried out by Ofqual in 2019 in response to the introduction of the current method of assessment (Ofqual, 2019).

A level science students answer questions worth up to 15% of the overall marks on their exams on the indirect assessment of practical skills (Abrahams et al, 2013). Further research could look at the success of students on these questions in comparison to those

that focus on wider theory content to understand in greater depth the relationship between hands-on practical work and success at these questions.

“For an assessment to be fair, the content, its context and the expectation of the assessment's performance should provide all candidates with an equal opportunity to demonstrate their ability.” (Isaacs et al, 2013) Further research could therefore look in more detail at the individual contexts of students during the pandemic and investigate the impact of that context on their practical skills in order to more fully understand how context did impact the fairness of this assessment.

Further research on Nisbet and Shaw’s typology of fairness (2020) could test the typology in other circumstances to investigate whether there are views about the distinctions between the senses of fairness to understand any other applications of the typology in greater depth.

Respondents to the survey were of the overwhelming opinion that the A level science Practical Endorsement is a superior method of assessment of practical skills in comparison to coursework. The assessment of speaking skills in A level modern foreign languages (MFL), is a situation where an endorsement might also be an appropriate method of assessment. To be prepared for future reforms to qualifications, further research could consider the application of the principles of the A level science endorsement in other subjects such as MFL.

References

Abrahams, I, Reiss, M. J. & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education, 49*(2), 209-251.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Baird, J.-A., Caro, D., Elliott, V., El Masri, Y., Ingram, J., Isaacs, T., Pinot de Moira, A., Randhawa, A., Stobart, G., Meadows, M., Morin, C. & Taylor, R. (2019). *Examination Reform: Impact of Linear and Modular Examinations at GCSE.* Oxford University Centre for Educational Assessment Report. OUCEA/19/2.

Baird, J-A, & Elliott, V. (2018). Metrics in education - control and corruption. *Oxford Review of Education, 44*(5), 533-544.

Caines, J., Bridglall, B. L. & Chatterji, M. (2014). Understanding validity and fairness issues in high-stakes individual testing situations. *Quality Assurance in Education, 22*(1), 5-18.

Cambridge Dictionary. (2022). https://dictionary.cambridge.org/dictionary/english/fairness

Camilli, G. (2006) Test Fairness. In R. Brennan (Ed.) *Educational measurement* (Fourth ed American Council on Education/Praeger series on higher education). Westport, Ct. pp. 221 – 256.

Carter, I. (2014). Is this the end of the English tradition of practical A-level science? *School Science Review, 96*(355), 12–14.

Childs, A. & Baird, J-A. (2020). General Certificate of Secondary Education (GCSE) and the assessment of science practical work: An historical review of assessment policy. *Curriculum Journal (London, England), 31*(3), 357-378.

Chamberlain, S. (2013). Qualification users' perceptions and experiences of assessment reliability. *Research Papers in Education, 28*(1), 118-133.

CLEAPSS. (2020). *Guide to doing practical work during the COVID-19 Pandemic – Science.* https://science.cleapss.org.uk/resource-info/gl343-guide-to-doing-practical-work-during-the-covid-19-pandemic-science.aspx

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth ed.). London, England; New York, New York.

Cramman, H., Kind, V., Lyth, A., Gray, H., Younger, K., Gemar, A., Eerola, P., Coe, R. & Kind, P., (2019) *Monitoring practical science in schools and colleges.* Durham University.

Department for Education (2010) *The Importance of Teaching: The Schools White Paper*. London, Department for Education.

Department for Education. UK. (2012). *Secretary of State for Education's letter to Ofqual.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/278935/2012-03-31-michael-gove-letter-to-glenys-stacey.pdf

Department for Education. UK. (2020). *Schools, colleges and early years settings to close.* https://www.gov.uk/government/news/schools-colleges-and-early-years-settings-to-close

Department for Education. UK. (2021) *Exams in 2022 – everything you need to know.* https://educationhub.blog.gov.uk/2021/11/11/exams-in-2022-everything-you-need-to-know/

El Masri, Y. H., Erduran, S., & Ioannidou, O. (2021). Designing practical science assessments in England: Students' engagement and perceptions. *Research in Science & Technological Education*, 1-21.

Erduran, S., El Masri, Y., Cullinane, A. & Ng, YPD. (2020). Assessment of practical science in high stakes examinations: A qualitative analysis of high performing English-speaking countries. *International Journal of Science Education, 42*(9), 1544-1567.

Evans, S. (2013). *A report on perceptions of current practical assessment in science GCSEs and IGCSEs.* Cambridge Assessment, Cambridge.

Field, A. P. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). Los Angeles; London.

Gatsby Charitable Foundation (2017). *Good Practical Science* [Online]. London, UK: The Gatsby Charitable Foundation. Available at: http://www.gatsby.org.uk/GoodPracticalScience.

He, Q., Boyle, A. & Opposs, D. (2011). Public perceptions of reliability in examination results in England. *Evaluation & Research in Education, 24*(4), 255-283.

Ipsos MORI. (2012). Fit for Purpose? *The view of the higher education sector, teachers, and employers on the suitability of A levels.* London: Ipsos MORI. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/377930/2012-04-03-fit-for-purpose-a-levels.pdf

Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Key concepts in Educational Assessment* (1st ed.). Sage.

Kane, M. (2010). Validity and fairness. *Language Testing, 27*(2), 177-182.

Moore, A., Fairhurst, P., Correia, C., Harrison, C., & Bennett, J. (2020). Science practical work in a COVID-19 world: are teacher demonstrations, videos and textbooks effective replacements for hands-on practical activities? *School Science Review, 102*(378), 7-12.

Nisbet, I. & Shaw, S. (2019). Fair assessment viewed through the lenses of measurement theory. *Assessment in Education: Principles, Policy & Practice, 26*(5), 612-629.

Nisbet, I. & Shaw, S. (2020). *Is assessment fair?* London.

Nisbet, I. & Shaw, S. (2022). Fair high-stakes assessment in the long shadow of Covid-19. *Assessment in Education: Principles, Policy & Practice,* published online.

Ofqual. (2013a). *Review of controlled assessment in GCSEs.* (Coventry, Ofqual).

Ofqual. (2013b). *Consultation on New A Level Regulatory Requirements.* (Coventry, Ofqual).

Ofqual. (2016). *GCE Subject Level Conditions and Requirements for Science (Biology, Chemistry, Physics) and Certificate Requirements*. May 2016. (Coventry, Ofqual).

Ofqual. (2017). *Comparable outcomes and new A levels*. March 2017. (Coventry, Ofqual). https://ofqual.blog.gov.uk/2017/03/10/comparable-outcomes-and-new-a-levels/

Ofqual. (2019). *The impact of qualification reform on the practical skills of A level science students. Paper 5: Final report on the pre- and post-reform evaluation of science practical skills.* (Coventry, Ofqual).

Ofqual. (2020a). *Consultation decisions - Proposed changes to the assessment of GCSEs, AS and A levels in 2021.* (Coventry, Ofqual).

Ofqual. (2020b). *Statement from Roger Taylor, Chair, Ofqual.* (Coventry, Ofqual). https://www.gov.uk/government/news/statement-from-roger-taylor-chair-ofqual

Ofqual. (2021a). *Consultation decisions - Proposed changes to the assessment of GCSEs, AS and A levels in 2022.* (Coventry, Ofqual).

Ofqual. (2021b). *Learning during the pandemic – quantifying lost time.* (Coventry, Ofqual).

Ofqual. (2021c). *Analysis of consultation responses - How GCSE, AS and A level grades should be awarded in summer 2021.* (Coventry, Ofqual).

Ofqual. (2021d). *Consultation on proposed changes to the assessment of GCSEs, AS and A levels in 2022 - Analysis of responses.* (Coventry, Ofqual).

Ofqual. (2021e). *Decisions on how GCSE, AS and A level grades will be determined in summer 2021.* (Coventry, Ofqual).

Ofqual. (2022). *Outcomes by Centre Type.* (Coventry, Ofqual).

Ofsted (2020a). *COVID-19 series: briefing on schools, October 2020. (*Manchester, Ofsted).

Ofsted (2020b). *COVID-19 series: briefing on schools, November 2020.* (Manchester, Ofsted).

Opposs, D. (2016). Whatever happened to school-based assessment in England's GCSEs and A levels? *Perspectives in Education, 34*(4), 52-61.

Qualifications and Curriculum Authority (QCA). (2005). *A review of GCE and GCSE coursework arrangements.* London: Qualifications and Curriculum Authority.

SCORE. (2009). *Assessment of practical work in science.* Summary of a seminar organised by SCORE at the Royal Society, 16th November 2009. Available at: https://www.stem.org.uk/system/files/elibrary-resources/2021/12/28409-assessment%20of%20practical%20work.pdf

Shaw, S. & Imam, H. (2013). Assessment of International Students Through the Medium of English: Ensuring Validity and Fairness in Content-Based Examinations. *Language Assessment Quarterly, 10*(4), 452-475.

Shaw, S. D., & Nisbet, I. (2021). Attitudes to fair assessment in the light of COVID-19. *Research Matters: A Cambridge Assessment publication, 31*, 6–21.

Sireci, S & Randall, J. (2021). Evolving Notions of Fairness in Testing. In B. Clauser & M. Bunch (Eds.), *The history of educational measurement: Key advancements in theory, policy, and practice* pp. 111-135). London.

Southall, E., Holmes, A., Hill, E. M., Atkins, B.D., Leng, T., Thompson, R. N., Dyson, L., Keeling, M.J., & Tildesley, M. J. (2021). An analysis of school absences in England during the COVID-19 pandemic. *BMC Medicine, 19*(1), 1-137.

Sutton Trust. (2022). *Comments on A level results day.* https://www.suttontrust.com/news-opinion/all-news-opinion/the-sutton-trust-comment-on-a-level-results-day/

Torrance, H. (2018). The Return to Final Paper Examining in English National Curriculum Assessment and School Examinations: Issues of Validity, Accountability and Politics. *British Journal of Educational Studies, 66*(1), 3-27.

Wade, N. & Abrahams, I. (2015) *Validity issues in the reform of a practical science assessment: an English case study.* Paper presented at 41st Annual Conference of the International Association for Educational Assessment, Kansas.

Wilson, F. (2013). *The assessment of practical science: a literature review.* Cambridge Assessment, Cambridge.

Wilson, F., Wade, N. & Evans, S. (2016). Impact of changes to practical assessment at GCSE and A-level: the start of a longitudinal study by OCR. *School Science Review, 98*(362), 119-128.

Worrell, F. (2016). Commentary on perspectives in fair assessment. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 283–294). New York, NY: Routledge.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147-170.

Zieky, M. J. (2016). Fairness in test design and development. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 9-31). New York, NY: Routledge.

Appendix A - Results from additional analysis for the one-way repeated measures ANOVA on hours of practical work pre-pandemic, in 2020-21 and in 2021-22

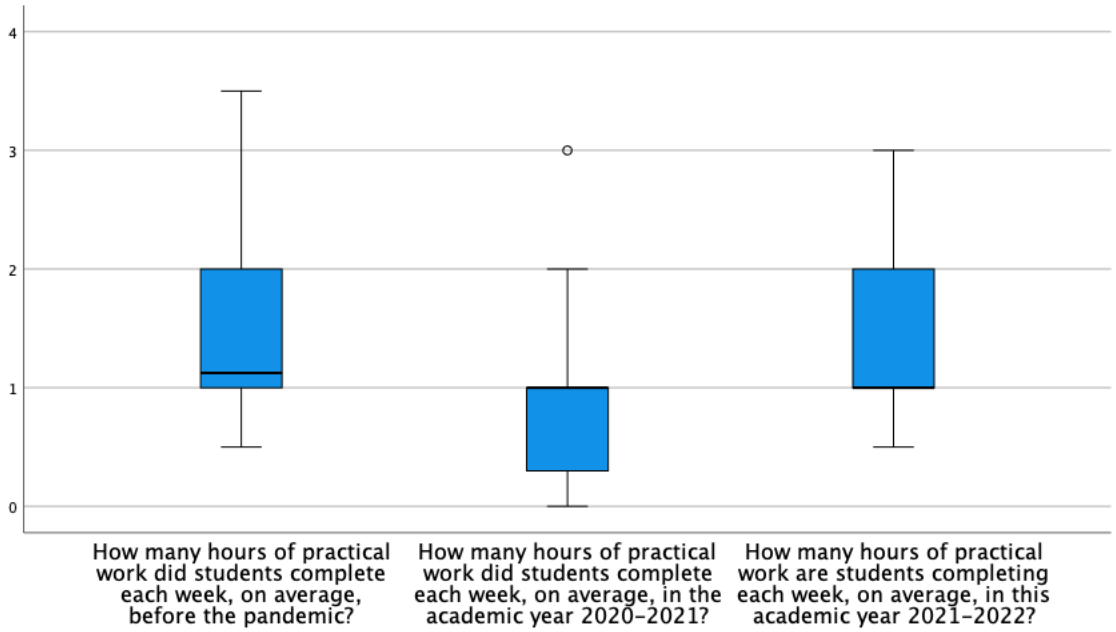


Figure A1. Boxplot showing number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22

A one-way repeated measures ANOVA was conducted on the number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 with the outlying data point removed.

The assumption of sphericity was not met, as assessed by Mauchly's test of sphericity, χ^2 (2) = 37.50, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied (ε = 0.68). The pandemic resulted in a significant reduction in the number of hours of practical work in 2020-21 (F(2,84) = 59.06, p = <0.001).

Post hoc analysis with a Bonferroni adjustment revealed that there was a reduction in the number of hours of practical work per week in 2020-21 (M = 0.77, SD = 0.54 hours) compared to the pre-pandemic number of hours (M = 1.48, SD = 0.73 hours), a significant mean reduction of 0.71 hours, 95% CI [0.49, 0.92], p = <0.001. There was also a reduction in the number of hours of practical work per week in 2020-21 (M = 0.77, SD = 0.54 hours) compared to the number of hours per week in 2021-22 (M = 1.45, SD = 0.71 hours), a significant mean reduction of 0.68 hours, 95% CI [0.47, 0.88], p = <0.001.

The hypothesis that there will be a significant reduction in the hours of practical work completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic is partially upheld. The difference in the number of hours carried out pre-pandemic (M = 1.48, SD = 0.73 hours) compared to 2021-22 (M = 1.45, SD = 0.71 hours) was not significant (0.03 hours, 95% CI [-0.07, 0.13], p = 1.00.

The raw data of number of hours of practical work in 2020-21 had a kurtosis value of 1.70, This raw data and that for the number of hours of practical work pre-pandemic and in 2021-22 was transformed to give a square root value of -1.08 for 2020-21 so a further one-way repeated measures ANOVA was conducted.

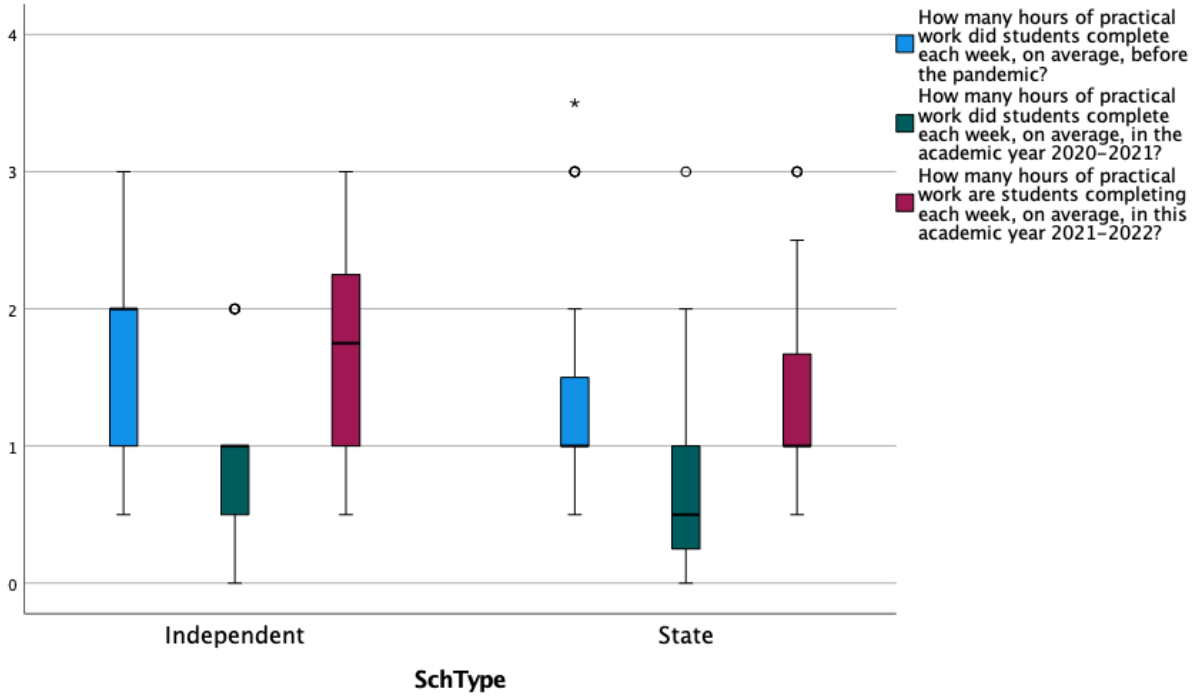
The assumption of sphericity was not met, as assessed by Mauchly's test of sphericity, χ^2 (2) = 37.19, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied (ε = 0.73). The pandemic resulted in a significant reduction in the number of hours of practical work in 2020-21 (F(2,85) = 41.45, p = <0.001).

Post hoc analysis with a Bonferroni adjustment revealed that there was a reduction in the number of hours of practical work per week in 2020-21 (M = 0.64, SD = 0.48 hours) compared to the pre-pandemic number of hours (M = 0.94, SD = 0.55 hours), a significant mean reduction of 0.30 hours, 95% CI [0.20, 0.39], p = <0.001. There was also a reduction in the number of hours of practical work per week in 2020-21 (M = 0.64, SD = 0.48 hours) compared to the number of hours per week in 2021-22 (M = 0.90, SD = 0.56 hours), a significant mean reduction of 0.26 hours, 95% CI [0.15, 0.36], p = <0.001.

The hypothesis that there will be a significant reduction in the hours of practical work completed by students per week in the 2021 and 2022 cohorts in comparison to before the pandemic is partially upheld. The difference in the number of hours carried out pre-pandemic (M = 0.94, SD = 0.55 hours) compared to 2021-22 (M = 0.90, SD = 0.56 hours) was not significant (0.04 hours, 95% CI [-0.20, 0.09], p = 0.34.

Appendix B - Results from additional analysis for the two-way mixed measures ANOVA on the interaction between hours of practical work pre-pandemic, in 2020-21 and 2021-22 and school type, independent or state

Figure B1. Boxplot showing number of hours of practical work pre-pandemic, in 2020-21



and in 2021-22 and school type, state and independent

A two-way mixed ANOVA was conducted on the number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent, with the outlying data points removed.

There was homogeneity of variances ($p > 0.05$) and covariances ($p = 0.01$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 22.36, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.74$).

There was no significant interaction between the number of hours of practical work carried per week out at the different points in time compared with independent and state school, $F(1, 52) = 0.05$, $p = 0.27$, partial $\eta^2 = 0.005$. Graph 1 below shows this lack of interaction.

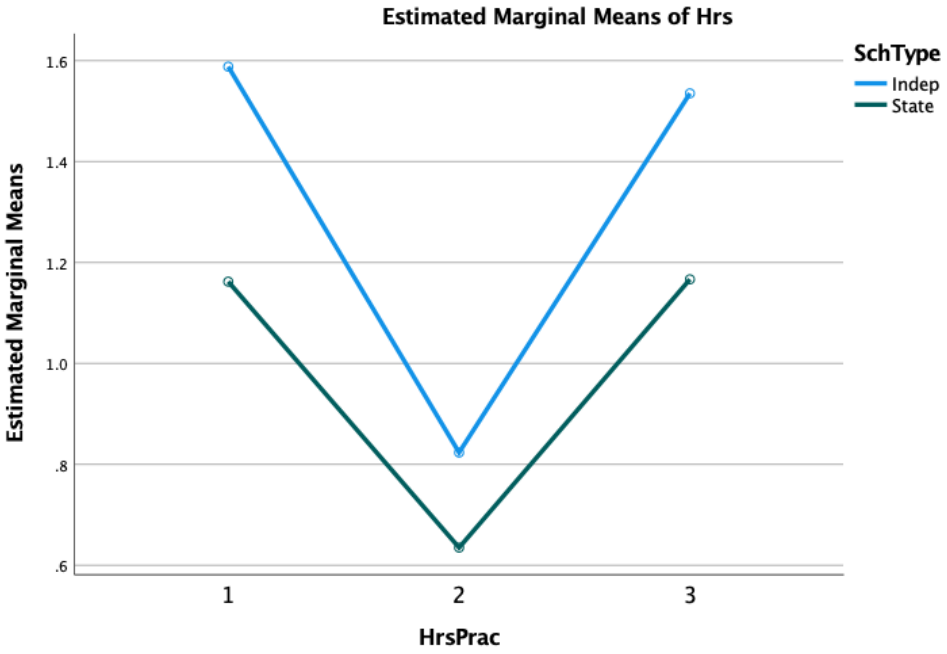


Figure B1. Graph showing the interaction between the mean hours of practical pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent with outliers removed

The raw data of the number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent, showed kurtosis as in table B1 below. The data was transformed to provide square root values which still showed kurtosis so the data was transformed to provide log 10 values as in table B2 below.

Table B1. Result of kurtosis calculation on data of number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent

School type	Kurtosis	
Independent	Pre-pandemic	-0.76
	2020-21	-0.32
	2021-22	-1.11
State	Pre-pandemic	1.34
	2020-21	4.36
	2021-22	1.19

Table B2. Result of kurtosis calculation on log 10 transformed data of number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent

School type	Kurtosis	
Independent	Pre-pandemic	-0.91
	2020-21	-0.21
	2021-22	-0.84
State	Pre-pandemic	-0.16
	2020-21	0.44
	2021-22	-0.27

A two-way mixed ANOVA was conducted on the number of hours of practical work pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent, using the log 10 transformed data.

There was homogeneity of variances ($p > 0.05$) and covariances ($p = 0.04$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 59.54, p = <0.001. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.60$).

There was no significant interaction between the number of hours of practical work carried per week out at the different points in time compared with independent and state school, $F(1, 54) = 4.23$, $p = 0.04$, partial $\eta^2 = 0.07$. Figure B2 below shows this lack of interaction.

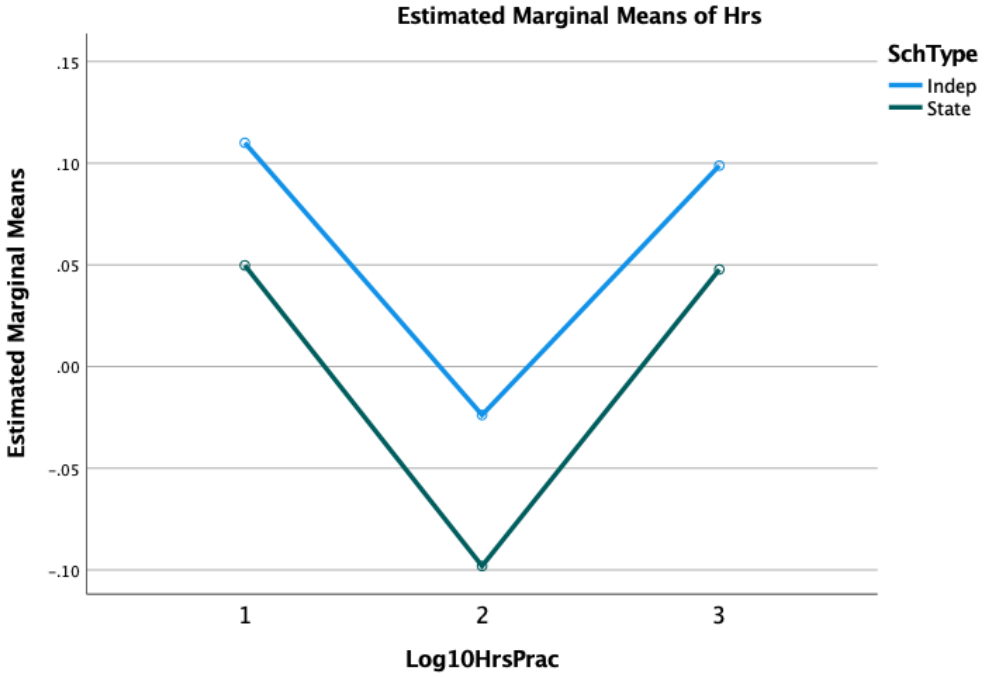


Figure B2. Graph showing the interaction between the mean hours of practical pre-pandemic, in 2020-21 and in 2021-22 and school type, state and independent using transformed log 10 data

Appendix C - Results from tests of assumptions of no outliers in data about the focus of practical work

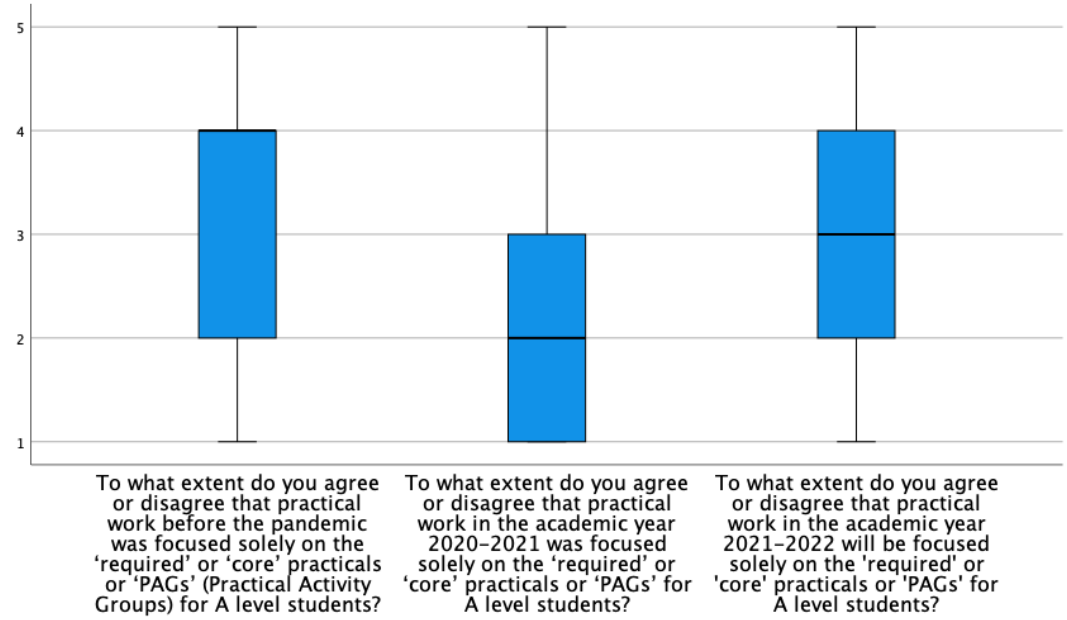


Figure C1. Graph showing the focus on 'required' or 'core practicals' or 'PAGs' for A level students, pre-pandemic, in 2020-21 and in 2021-22 as reported by teachers

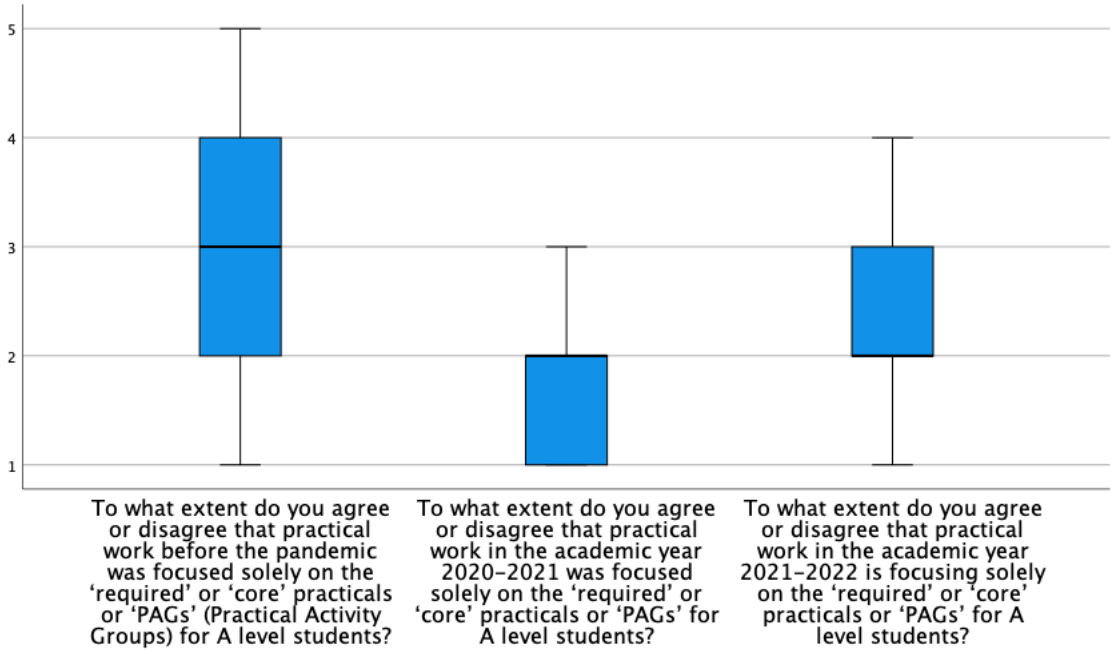


Figure C2. Graph showing the focus on 'required' or 'core practicals' or 'PAGs' for A level students, pre-pandemic, in 2020-21 and in 2021-22 as reported by monitors

Appendix D – Results from analysis to show extreme kurtosis on fairness scales – raw data, transformed to the square root and transformed to the log10

Descriptive statistics	Fairness type – raw data					
	Implied Contractual	Retributive	Consequential	Retrospective	Formal	Relational
N	81	81	81	81	81	81
Mean	2.98	2.91	3.04	3.29	2.92	2.70
Std. Deviation	0.94	0.93	0.83	1.10	0.92	0.86
Kurtosis	1.36	2.10	3.16	-0.24	1.12	2.60

Table D1. Result of kurtosis calculation on fairness scales – raw data

Descriptive statistics	Fairness type – square root					
	Implied Contractual	Retributive	Consequential	Retrospective	Formal	Relational
N	81	81	81	81	81	81
Mean	1.69	1.66	1.71	1.78	1.67	1.67
Std. Deviation	0.36	0.39	0.33	0.36	0.36	0.36
Kurtosis	10.56	10.35	15.65	6.47	10.26	10.26

Table D2. Result of kurtosis calculation on fairness scales – square root values

Descriptive statistics	Fairness type – log 10					
	Implied Contractual	Retributive	Consequential	Retrospective	Formal	Relational
N	81	81	81	81	81	81
Mean	0.23	0.22	0.24	0.25	0.22	0.22
Std. Deviation	0.07	0.07	0.06	0.08	0.07	0.07
Kurtosis	2.04	2.21	3.90	0.76	1.48	1.48

Table D3. Result of kurtosis calculation on fairness scales – log 10 values

Appendix E - Exploratory factor analysis on fairness items

Fairness	Item	Factor					
		1	2	3	4	5	6
Formal	The rules in place for the assessment of the CPAC were not always appropriately followed	0.72			-0.44		
	The rules in place for the assessment of the CPAC were sometimes broken	0.74			-0.41		
	It was not always possible to follow the rules in place for the assessment of the CPAC	0.62	-0.36				
	It was not possible to distinguish between students with weak or strong practical skills as fewer than 12 practical activities could be used to assess the CPAC	0.66					0.31
Implied contractual	The assessment of the CPAC was not carried out as students and teachers had legitimately expected	0.74					
	Teachers legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic	0.67		0.35			
	Students legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic	0.80					
Relational	Not all students were treated the same way in the assessment of the CPAC	0.71			-0.46		
	Groups of students were treated differently to each other in the assessment of the CPAC	0.57	0.35		-0.30		
	Even prior to the pandemic, students have never all consistently had the same access to technician support when being assessed on the CPAC	0.56	0.51				
	Even prior to the pandemic, students have never all consistently had the same access to specialist teachers when being assessed on the CPAC	0.43	0.61	-0.43			
	Even prior to the pandemic, students have never all consistently had the same access to teachers that are confident and experienced when being assessed on the CPAC	0.61	0.49				
	Even prior to the pandemic, students have never had the same access to practical resources when being assessed on the CPAC	0.63					

	Students could have been assessed on a number of practical skills that varied widely	0.60				-0.50	
	Students could have been assessed on types of practical skills that varied widely	0.69				-0.38	
	This cohort of students were not treated in the same way as previous cohorts of students in the assessment of the CPAC	0.57		0.48			0.38
Retributive	It was not always possible to appropriately reward students for their competence in their practical skills	0.56		0.34	0.39		
	The result of the endorsement did not always reflect student competence in practical skills	0.68					
	The assessment of the CPAC did not always reward students for their competence in practical skills	0.56		-0.49			
Consequential	Students will be treated unjustly in the future based on their result in the Practical Endorsement	0.73				0.33	
	Students will be treated unjustly in the future based on their practical skills as assessed by the CPAC	0.70	-0.30				
	Students with weak practical skills could use their pass in the Practical Endorsement to access university courses where this is a requirement	0.50		0.35		0.30	
	Students with weak practical skills will have opportunities that they would not have had if their practical skills had been assessed as required prior to the pandemic	0.51		0.38			
	Students with strong practical skills will not have the opportunities they would have had if all students had been assessed as required prior to the pandemic	0.62					0.36
Retrospective	The assessment of the CPAC has never been a fair assessment method of practical skills	0.62	-0.40				0.32
	The award of a pass in the Practical Endorsement has never been a fair reflection of competence in the skills assessed by the CPAC	0.70	-0.45				

Appendix F – CUREC approval

[redacted]
Date: Thursday, 3 February 2022 at 18:17
[redacted]
Subject: CUREC Ref: CIA-22-085

Dear [redacted]

Title: The impact of the COVID-19 pandemic on the fairness of the direct assessment of practical work in the A level sciences

Ref: CIA-22-085

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Best wishes

Member of the DREC

Appendix G – Survey questions

1. Over 18
2. Consent to take part

Job / Role

3. Which best describes you? Teacher/Exam board monitor/Teacher and exam board monitor

TEACHERS

4. Which best describes your school or college? 11-19 maintained school/11-19 academy/11-19 independent/ further education college/sixth form college/11-19 special school/other
5. Which A level science subject do you teach? If you teach more than one subject, please select one as the focus of this survey. Biology/Chemistry/Physics/Geology
6. How many students are typically entered for the full A level in your subject?
7. How many years have you been teaching A level students?

Frequency/number/focus of practical activities

Before the pandemic

8. How many hours of practical work did students complete each week, on average, before the pandemic?
9. How many practical activities did you use, on average, to assess the Common Practical Assessment Criteria (CPAC) across the full A level course before the pandemic?
10. To what extent do you agree or disagree that practical work before the pandemic was focused solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD[3]

Students entered for A levels in 2021 (2021 cohort)

11. How many hours of practical work did students complete each week, om average, in the academic year 2020-2021?
12. Did you assess the CPAC on the basis of fewer than the required 12 practical activities for the 2021 cohort? Y/N
13. If yes, how many practical activities did you use, on average, to assess the CPAC for the 2021 cohort?
14. To what extent do you agree or disagree that practical work in the academic year 2020-2021 was focused solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD
15. To what extent do you agree or disagree that students in the 2021 cohort in your school or college did the same number of practical activities **as each other** when being assessed on the CPAC? SA/A/NA or D/D/SD
16. If you disagreed or strongly disagreed to the previous question, please explain how the number varied and why?

Students entered for A levels in 2022 (2022 cohort)

17. How many hours of practical work are students completing each week, on average, in this academic year 2021-2022?

[3] Strongly agree/agree/neither agree or disagree/disagree/strongly disagree

18. Will you assess the CPAC on the basis of fewer than the required 12 practical for the 2022 cohort? Y/N
19. If yes, how many practical activities will you use, on average, to assess the CPAC for the 2022 cohort?
20. To what extent do you agree or disagree that practical work in the academic year 2021-2022 will be focused solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD
21. To what extent do you agree or disagree that students in the 2022 cohort in your school or college will do the same number of practical activities **as each other** when being assessed on the CPAC? SA/A/NA or D/D/SD
22. If you disagreed or strongly disagreed to the previous question, please explain how the number will vary and why?

<u>Type of practical activities</u>

Before the pandemic

23. What types of practical work did A level students carry out in your school or college before the pandemic? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations
    - Teacher demonstrations
    - Simulations
    - Fieldwork
    - Other
24. If you answered 'other' to the previous question, please specify.

Students entered for A levels in 2021 (2021 cohort)

25. What types of practical work did A level students carry out in your school or college in the academic year 2020-2021? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations
    - Teacher demonstrations
    - Simulations
    - Fieldwork
    - Other
26. If you answered 'other' to the previous question, please specify.
27. To what extent do you agree or disagree that students in the 2021 cohort in your school or college did the same type of practical activities **as each other** when being assessed on the CPAC? SA/A/NA or D/D/SD
28. If you disagreed or strongly disagreed to the previous question, please explain how the type varied and why?

Students entered for A levels in 2022 (2022 cohort)

29. What types of practical work are A level students carrying out in your school or college in the academic year 2021-2022? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations

- Teacher demonstrations
- Simulations
- Fieldwork
- Other

30. If you answered 'other' to the previous question, please specify.
31. To what extent do you agree or disagree that students in the 2022 cohort in your school or college will do the same type of practical activities as each other when being assessed on the CPAC? SA/A/NA or D/D/SD
32. If you disagreed or strongly disagreed to the previous question, please explain how the type will vary and why?

ROUTING

33. Please choose the correct route through the rest of the survey.

EXAM BOARD MONITORS

34. Which A level science subject do you usually monitor? Biology/Chemistry/Physics/Geology
35. How long have you been carrying out monitoring visits?
36. Which exam board do you work for? AQA/Eduqas/OCR/Pearson

Please use your overall experience of monitoring the A level sciences to answer the following questions.

Frequency/number/focus of practical activities

Before the pandemic

37. How many practical activities, on average, did teachers use to assess the Common Practical Assessment Criteria (CPAC) before the pandemic?
38. To what extent do you agree or disagree that practical work before the pandemic was focused solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD

Students entered for A levels in 2021 (2021 cohort)

39. In general, did teachers use fewer than the required 12 practical activities to assess the 2021 cohort in the CPAC? Y/N
40. If yes, how many practical activities, on average, were used to assess the 2021 cohort in the CPAC?
41. To what extent do you agree or disagree that practical work in the academic year 2020-2021 was focused solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD

Students entered for A levels in 2022 (2022 cohort)

42. In general, are teachers using fewer than the required 12 practical activities to assess the 2022 cohort in the CPAC? Y/N
43. If yes, how many practical activities, on average, are being used to assess the 2022 cohort in the CPAC?
44. To what extent do you agree or disagree that practical work in the academic year 2021-2022 is focusing solely on the 'required' or 'core' practicals or 'PAGs' for A level students? SA/A/NA or D/D/SD

Type of practical activities

Before the pandemic

45. Based on your monitoring visits, what types of practical work did A level students carry out before the pandemic? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations
    - Teacher demonstrations
    - Simulations
    - Fieldwork
    - Other
46. If you answered 'other' to the previous question, please specify.

Students entered for A levels in 2021 (2021 cohort)

47. Based on your monitoring visits, what types of practical work did A level students carry out in the academic year 2020-2021? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations
    - Teacher demonstrations
    - Simulations
    - Fieldwork
    - Other
48. If you answered 'other' to the previous question, please specify.

Students entered for A levels in 2022 (2022 cohort)

49. Based on your monitoring visits, what types of practical work are A level students carrying out in the academic year 2021-2022? Tick all that apply
    - Experiments following a method
    - Experiments involving some element of investigation
    - Whole open-ended investigations
    - Teacher demonstrations
    - Simulations
    - Fieldwork
    - Other
50. If you answered 'other' to the previous question, please specify.

TEACHERS AND MONITORS

Fairness

51. Were all students in your school or college treated in the same way when being assessed on the CPAC?
52. Please explain your answer to the previous question.
53. Were all students across different schools and colleges treated in the same way when being assessed on the CPAC?
54. Please explain your answer to the previous question.

To what extent do you agree or disagree that during the pandemic, fairness may have been reduced because - SA/A/NA or D/D/SD

55. The rules in place for the assessment of the CPAC were not always appropriately followed.
56. The rules in place for the assessment of the CPAC were sometimes broken.
57. It was not always possible to follow the rules in place for the assessment of the CPAC.
58. It was not always possible to appropriately reward students for their competence in their practical skills.
59. The result of the endorsement did not always reflect student competence in practical skills.
60. The assessment of the CPAC did not always reward students for their competence in practical skills.
61. It was not possible to distinguish between students with weak or strong practical skills as fewer than 12 practical activities could be used to assess the CPAC.
62. Students could have been assessed on a number of practical skills that varied widely.
63. Students could have been assessed on types of practical skills that varied widely.
64. The assessment of the CPAC was not carried out as students and teachers had legitimately expected.
65. Teachers legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic.
66. Students legitimately expected the assessment of the CPAC to happen in a way that was then not possible due to the pandemic.
67. Not all students were treated the same way in the assessment of the CPAC.
68. This cohort of students were not treated in the same way as previous cohorts of students in the assessment of the CPAC.
69. Groups of students were treated differently to each other in the assessment of the CPAC.
70. Students will be treated unjustly in the future based on their result in the Practical Endorsement.
71. Students will be treated unjustly in the future based on their practical skills as assessed by the CPAC.
72. Students with weak practical skills will have opportunities that they would not have had if their practical skills had been assessed as required prior to the pandemic.
73. Students with weak practical skills could use their pass in the Practical Endorsement to access university courses where this is a requirement.
74. Students with strong practical skills will not have the opportunities they would have had if all students had been assessed as required prior to the pandemic.
75. The assessment of the CPAC has never been a fair assessment method of practical skills.
76. The award of a pass in the Practical Endorsement has never been a fair reflection of competence in the skills assessed by the CPAC.
77. Even prior to the pandemic, students have never all consistently had the same access to technician support when being assessed on the CPAC.
78. Even prior to the pandemic, students have never all consistently had the same access to specialist teachers when being assessed on the CPAC.
79. Even prior to the pandemic, students have never all consistently had the same access to teachers that are confident and experienced when being assessed on the CPAC.

80. Even prior to the pandemic, students have never had the same access to practical resources when being assessed on the CPAC.
81. In light of the questions above about fairness, in what ways do you think the assessment of the CPAC is fair or unfair as a result of the pandemic? Free response, word limited.

INTERVIEW

Finally, if you are interested in contributing further to this research on the fairness of the award of the A level sciences Practical Endorsement and would be willing to be interviewed via Zoom or MS Teams to explore your answers in more detail, please provide your name and email address below.

82. Name
83. Email address

Appendix H – Interview questions

Thank you very much for completing the survey and for agreeing to this interview. I just thought it would be helpful to share the aims of the research and what I would like to ask you about.

I am essentially interested in 2 things – what it has been like in the classroom over the last couple of years – the challenges for practical work and the assessment of practical skills, whether students have had access to the same amount of practical work as before and now, as the impacts of the pandemic hopefully lessen. And also, what impact this has had on the fairness of the award of the endorsement. I based the questions in the survey on fairness on a typology from a recently published book called "Is assessment fair?". The typology describes several ways to approach fairness in assessment and I am interested in the application of these interpretations to the assessment of practical skills over the last couple of years.

I have a few questions to guide our discussion, but I don't want to stick to rigid questions and answers – more let's just have a chat and I may well use the questions to delve a bit deeper into anything you raise. And, just for your information, Zoom will generate a transcription of our conversation, so I don't need to record the interview.

1. So, to start, please can you tell me about what the delivery and assessment of practical skills has been like over the last couple of years. In your experience, how did the pandemic affect this and what were the challenges?

2. In terms of the rules in place for the assessment of the CPAC, we, as in Ofqual, relaxed them slightly for 2021 and 2022 – we were aware that it may have been difficult for a variety of public health reasons for students to do the number of practicals needed for the endorsement – 12 – and so said that as long as enough practical work was done to allow all of the criteria to be assessed, teachers could do fewer than 12. For this summer, of course, there needed to be consideration of the 15% of marks assigned to the assessment of practical skills in the written exams. So, what is your view of the rules in this context?

   Follow up – what do you think about how the rules may have impacted fairness? For example, normally we can be confident that a pass in the endorsement reflects the completion and assessment of at least 12 practicals but that may not have been the case over the last couple of years? Might it have varied hugely from one school to another? Might this have been more variation than normal?

3. So, teachers and students will have started A level science courses with a reasonable assumption that practical work will be an important part of the course. Do you think the impact of the pandemic on practical work allowed those expectations to be met?

   Follow up – do you think that fairness was affected in light of your comments about these expectations?

4. Thinking about students and their experience of practical work over the last couple of years, what is your view on their likely experience?

   Follow up - do you think they were treated fairly in terms of a comparison of the experience in one school with that in another school?

5. So, since the endorsement was first awarded in 2017, teachers, students and indeed HE have had a broad understanding as to what a pass means in terms of practical competence e.g., at least 12 practicals, assessment against the CPAC, practicals that are designed to cover the apparatus and techniques in the specification. What is your view about whether the results awarded in 2021 and 2022 will mean the same between one school and another and in comparison to those awarded in the past and in the future?

   Follow up – do you think there is any issue of fairness based on your comments?

6. Some HE courses require a pass in the endorsement and it is part of the offer that students get from the university. Thinking about the issues of whether this result means the same in a broad sense, what do you think about how students might use their endorsement result?

   Follow up – might the results be used in an unfair way in the future? And if so, unfair to who?

7. And as a final question, is there anything else about fairness and the assessment and award of the Practical Endorsement that you would like to tell me about, that the questions haven't addressed?