Overview of the Perceptions of Assessment Standards in Scotland Project

Report 1 of 4

Jo-Anne Baird, Thomas Godfrey-Faussett, Louise Hayward, Carolyn, Hutchinson, Ashmita Randhawa, Ernest Spencer, Lesley Wiseman-Orr and SQA colleagues

September 2022

Contents

Perceptions of Assessment Standards in Scotland Project

Understanding stakeholder views of qualifications in Scotland

The research reported here is part of a programme of work by SQA to engage with stakeholders' perceptions of qualifications in Scotland. Prompted by the systems crisis in the summer of 2020 generated by the pandemic, the impetus for this research programme came from the SQA Executive, which charged the research team and the organisation to investigate what qualification users considered the principal issues to be and to use the information to better develop assessment policy. Dips in public confidence occur following systemic crises and it takes several years to regain trust and confidence. The Scottish Government initiated a rapid review of the 2020 National Qualifications experience. In the review report, attention was drawn to a perceived lack of transparency in engagement with stakeholders (Priestley et al., 2020, p4) and this project formed part of the response to these issues.

This report is the first in a series of four reports which look at stakeholder perceptions of standards, in this case giving an overview of the project and its findings. The other three in this series are:

2. Perceptions of Assessment Standards in Scotland: Focus Groups with Stakeholders
3. Perceptions of Assessment Standards in Scotland: Questionnaire with Stakeholders
4. Perceptions of Assessment Standards in Scotland: Employer Survey

This report draws together findings on the main themes across the three sets of data collection.

Following the public reaction to grading of the summer 2020 National Qualifications in Scotland, SQA launched a broad research programme to investigate and engage with stakeholders' views regarding qualifications in Scotland. This project forms part of the programme of work. The Perceptions of Assessment Standards (PASS) in Scotland project specifically investigated stakeholders' views on fairness and standards of qualifications in 2020 and 2021. An empirical study of the perceptions of stakeholders and their reports on what influenced their views was conducted through focus groups with 82 stakeholders, a questionnaire with 918 stakeholders and a telephone survey with 103 employers. The aims of the work were to understand views of qualifications in Scottish society and to use this to inform better assessment policymaking and communication with stakeholders.

The research was conducted by teams from the University of Glasgow, the University of Oxford and SQA. The current report integrates the findings thematically across the three methods of data collection and links them with the broader research literature, to draw conclusions about views regarding standards and fairness of qualifications in Scotland in 2020 and 2021. Over the period of the research, several policy reviews on national assessment were undertaken on behalf of the Scottish Government. Public debates relating to their conclusions and proposals are likely to have influenced the findings of this research.

Stakeholders' perceptions of assessment standards in Scotland

Public confidence in any qualification system is essential to maintain the currency of the certificates for the learners whose life chances depend upon them. It is vital that relevant stakeholder groups are both well-informed and consulted about these systems. Surprisingly, how people view standards and what influences their confidence in assessment systems are little-researched topics. We know from research conducted in England that stakeholders' views of reliability and trust in the system go well beyond the standard setting system, to include issues regarding content standards, bias, reliability, validity, and utility of qualifications (He, Boyle and Opposs, 2011; Chamberlain, 2013; Simpson and Baird, 2013). Necessarily, assessment industry insiders have come to define terms in specific ways, but this can make communication between assessment specialists and the wider community less effective.

The purpose of this research was to investigate what standards mean to practitioners, pupils, parents and to society more broadly. Influences upon stakeholders' confidence in qualifications and their views about assessment reform were also matters of interest. Additionally, the disruptive context of the pandemic raised significant issues upon which stakeholders' views were sought. Concerns regarding assessment standards during the pandemic were an important part of the basis for the evidence-gathering. As this is the first empirical work on stakeholders' views of the meaning of assessment standards in Scotland, we do not know how the pandemic has affected the data we gathered, nor are we able to compare views before and during the pandemic. The emerging findings informed SQA's Communications Strategy and through early dissemination within SQA, contributed to the delivery of that strategy.

Approach to the research

Research, which was at once independent, well-informed, and integrated with the operational needs of SQA was required. Additionally, the project needed to gather information that could influence longer term assessment policy as well as the shorter-term needs for forthcoming qualifications diet. A participatory research design was adopted, involving collaboration between researchers in the Departments of Education at the Universities of Glasgow and Oxford and colleagues in SQA's research and communications teams. Stakeholders were consulted about the design of the project and the data collection techniques. To ensure that the project was open to the views of participants, focus groups were first conducted with 82 practitioners (teachers, lecturers and school and college managers), pupils, students, parents, carers, employers, university admissions officers and journalists. Next, an online questionnaire survey was conducted with 918 complete responses. Additionally, a telephone survey of 103 employers was conducted.

As the focus groups were conducted in advance of the certification of the 2021 National Qualifications, rapid analyses of the findings by stakeholder group were produced for use by relevant SQA colleagues. Questions were raised by teams across SQA that were addressed in the data collection where possible. Interpretation of the findings benefited from wide discussion across the project team and with colleagues across SQA, but importantly, the voices of stakeholders were deemed central to the purpose of the work throughout. More detail about the methodology of each of the data collection techniques can be found in the other three reports in this series.

This study was conducted in accordance with the guidelines from the British Educational Research Association (BERA). Research ethics approvals were obtained for the study ahead of commencing data collection through the University of Oxford (ED-CIA-21-157) and University of Glasgow (400200142) ethical approval procedures.

Effects of the pandemic on education and assessment in Scotland

The impact of COVID-19-19 was felt globally, with 1.5 billion school-age pupils across 188 countries estimated to be out of school in 2020 (OECD, 2021a). A normal school year has approximately 190 days and between March 2020 and April 2021, schools were closed for 119 of those days in Scotland (Major, Eyles and Machin, 2021). There was remote schooling during this period, for which provision varied across the country. On average, 64 days of schooling were lost to COVID-19 in Scotland once this is taken into account. This compared with 61 days in England and Northern Ireland and 66 days in Wales (Major, Eyles and Machin, 2021). As well as learning loss (EEF, 2021), there were effects upon the health and wellbeing of children and young people, especially their mental health and social confidence (Education Scotland, 2021). Access to remote education was affected by the digital divide; the most disadvantaged were less able to access technology, the internet, and a quiet place to learn or they had family responsibilities that kept them from their studies (Education Scotland, 2021).

Holding examinations in highly populated school halls was a health risk under the pandemic and countries adapted their assessment systems in five ways (Opposs, 2020):

1. Cancellation of final examinations - use of completed examinations and any school-based assessments
2. Cancellation of final examinations – use of school- and college-generated estimates
3. Replacement of final examinations with online assessments taken at home
4. Postponement of final examinations
5. Holding examinations on schedule.

Countries such as Finland, Germany and Poland decided to hold the examinations on schedule, whereas China, Australia, Spain, and several other countries postponed the examinations. In fact, postponement was the prevalent strategy (UNESCO, 2020). Across the British Isles, statistical moderation of teacher estimates of grades was used in 2020. These systems varied in each jurisdiction and the arrangements in Scotland are explained in more detail next.

Arrangements for National Qualifications in Scotland in 2020 and 2021

Assessment arrangements in academic sessions 2019-20 and 2020-21 were badly disrupted by the impacts of the COVID-19-19 pandemic. The specific circumstances of each year's disruption meant that different approaches to the assessment of learners were adopted, with 2020 grading for National Qualifications based on teachers' and lecturers' judgments, and 2021 grading based on demonstrated attainment, with teachers and lecturers assessing internally produced evidence.

In 2020, the first national lockdown took place from 23 March. The Scottish Government announced on 19 March that all schools were to close from that date and that the examination diet, which generally runs from the end of April until the beginning of June, would be cancelled. This also affected the uplift of coursework that would usually be marked or verified by SQA and meant that no assessment evidence could be used in the normal way. The model adopted by SQA to provide grades and awards involved four stages (SQA, 2020a). At the first stage, teachers and lecturers were, after online training, asked to submit estimated banded grades for, and a rank ordering of, the learners in each subject at National 5, Higher and Advanced Higher in their school, college, or other type of centre. Quantitative and qualitative data were then used to moderate these estimates, based partly on centres' previous patterns of estimates and results (SQA, 2020b). Learners were issued with these moderated results on 4 August 2020. There was a fourth stage of the process, where centres could appeal by submitting evidence to SQA for any learner whose moderated grade was lower than their estimated grade. However, this final stage was abandoned due to public outcry over the moderation process, and particularly its use of statistical methods. On 11 August, Scottish Government announced that they were directing SQA to recertificate learners using their estimated grades, except where moderation had increased their grade, and these final results would be published on 21 August 2020 (Scottish Government, 2020a).

Following a national consultation, SQA modified course assessments for 2020-21 to make space for more learning and teaching time and to take account of COVID-19-related public health advice (SQA, 2020c). This varied by subject but included measures like removing coursework requirements, removing assessments not possible due to public health guidance, and giving guidance on which topics would be assessed. An Alternative Certification Model (ACM) was also developed by SQA and other stakeholders across Scottish education, through the National Qualifications 2021 group. This was intended as a contingency if exams were cancelled for a second time, and this cancellation for first National 5 and then Higher and Advanced Higher was announced in late 2020.

The 2021 ACM approach relied on demonstrated attainment (SQA, 2021a). SQA provided subject-specific guidance and assessments and schools, colleges, and other centres, alongside their local authorities, could decide whether and how to use SQA's assessment materials, or whether to adopt a different approach to assessment. Teachers and lecturers then used their professional judgement to determine provisional grades for their learners based on the assessment evidence available. The grading was moderated internally and within local authorities, with SQA conducting a quality assurance process through sampling of the evidence. Learners were certified with these results on 10 August 2021, and for the first time could appeal directly to SQA if they were dissatisfied (SQA, 2021b).

The policy context

The PASS project was undertaken in a frenetic policy context. The COVID-19 pandemic had destabilised society, including the education system, and National Qualifications. The decision to cancel the examination diet from 2020 was announced by the Deputy First Minister, John Swinney, on the 19th of March 2020. In his speech to the Scottish Parliament, he acknowledged the exceptional nature of that decision,

> *In all of our history, Scotland has never cancelled the exams. Since 1888, they have been held every May or June without fail. In the midst of two world wars, the exams went ahead. It is a measure of the gravity of the challenge we now face, that I must today announce the exams will not go ahead this year.*

The uncertainty of whether the examinations would be able to go ahead had been under consideration for some time before the announcement and the SQA had to respond to change in unprecedented ways, planning for a range of potential scenarios in a timescale that would have been considered impossible before the pandemic.

At the same time, the process of evidence collection was underway for two related OECD visits. The original plan of Scottish Government had been to review the Senior Phase of Curriculum for Excellence (16-18), but a defeat in parliament in January 2020, led the Deputy First Minister to announce that OECD had been invited to undertake a full review of Curriculum for Excellence (CfE) including the Broad General Education (3-15) and how that articulated with the senior phase. The OECD review of progress made in CfE since 2015 took place in the time of COVID-19. Public Health restrictions meant that both OECD 'visits' to Scotland were virtual; the first from 28 September-2 October 2020 and the second 2-5 November 2020. The first OECD report *Scotland's Curriculum for Excellence: into the future*, was published in June 2021. Publicity associated with the publication of the first OECD report made little reference to the positive statements made by OECD. In the foreword, Andreas Schleicher indicated,

> *The OECD team finds that CfE continues to offer a vision and a philosophy of education widely supported and worth pursuing. ... Just as Scotland's Curriculum for Excellence was among the pioneers of 21st century learning, its most recent developments hold valuable lessons for other education systems and their own curriculum policies.*

<div align="right">OECD (2021b, Foreword)</div>

The report set out 12 recommendations linked to a series of proposed actions designed to strengthen CfE and to tackle implementation challenges. Two recommendations are of particular significance for this study. The first related to a perception of a lack of alignment between the philosophy of CfE and the Senior Phase

> *Recommendation 1.3. Adapt the Senior Phase to match the vision of CfE*
>
> *Scotland could consider adapting the pedagogical and assessment practices and the structure of learning pathways in the Senior Phase to enhance learners' experience of upper-secondary education and help them develop CfE's four capacities continuously.*

The second related to a perceived lack of alignment between curriculum, qualifications, and evaluation:

'Recommendation 3.3. Align curriculum, qualifications, and system evaluation to deliver on the commitment of Building the Curriculum 5

Scotland could first identify modes of student assessment that could be used in school and external settings at Senior Phase levels, in alignment with the four capacities and CfE philosophy; and second, re-develop a sample-based evaluation system to collect robust and reliable data necessary to support curriculum reviews and decision making.

The second OECD report focused on qualifications in the senior phase. Led by Professor Gordon Stobart, this review concluded that although CfE was regarded as a pioneering example of curriculum reform, the senior phase examinations were perceived to have 'seen less reform.' Stobart argued that the cancellation of examinations in 2020 and 2021 offered an opportunity to 'radically consider the assessment system'. The review compared the system in Scotland with four other British legacy systems and five other qualifications traditions to provide an evidence base on which Scotland might reflect as it considered how alignment might be enhanced between CfE and assessment in the senior phase. The report identified three major themes. In Scotland,

- external assessments could be more innovative to capture a wider range of student capabilities;
- the role of teacher assessment could be reconsidered; and
- the academic and vocational strands could be better integrated with the assessment system to offer a broader range of curriculum options.

In June 2021, the Cabinet Secretary accepted all 12 recommendations and set out a plan for taking the recommendations forward. As part of this response, the Cabinet Secretary announced the intention to replace the Scottish Qualifications Authority and to reform Education Scotland and established a review led by Professor Ken Muir to offer advice on the future configuration of these two national agencies.

Acknowledgements

Findings of the PASS project

Here, we present the main findings from the project, building upon previously published research. From the research literature outlined below, we can conclude that transparency and trust have a complex relationship and that communication regarding qualifications needs careful management. Media agendas may be counterproductive to trust in qualifications and stakeholders may have quite different concerns to awarding body staff.

Teacher assessments are typically graded more generously than external examinations (Urhahne and Wijnia, 2021). Whilst there is evidence for bias in teacher assessments in some of the previous research literature, reassuringly, the results in England in 2020 (where there was evidence) did not generally exacerbate attainment gaps that existed in normal years.

Recent work on notions of fairness during the pandemic concluded that, even prior to the pandemic, the assessment industry had not taken enough account of fairness at the level of the individual. Relational fairness, in which individuals complained that the assessment results where unfair if other pupils like them were treated differently, came to the fore. Additionally, retributive fairness, in which the assessment results reflect pupils' just awards, was a feature of the prevailing narrative. This reflects reported concerns surrounding the perceived learning loss during the pandemic. Our research and that of others signalled that grade inflation fairness concerns receded, but it remains to be seen whether these will return in public debates. Different definitions of assessment standards are used in qualifications. The only approach to defining standards which appears to be consistent with how standards were set in Scotland and in other parts of the UK for national qualifications during the pandemic appears to be 'criterion referencing.' We deal with these topics in more detail below, first outlining the previous research literature and then turning to our findings. Finally, we took the opportunity to explore stakeholders' views regarding the future of assessment in Scotland and the lessons that could be learned from the pandemic and outline our findings on these issues below. We present a thematic synthesis of the results across the three methods of data collection: focus groups with stakeholders (report 2), questionnaire with stakeholders (report 3) and survey of employers' views (report 4).

Standards

In keeping with the other nations of the UK, assessment outcomes were higher in Scotland in 2020 and 2021 compared with previous years. In the TeenCOVID-19Life survey of students who received qualification results in 2020, half of them thought their grades would have been the same if they had sat examinations that year, approximately one third believed that they would have got higher grades and 15% thought their grades would have been lower (TeenCOVID-19Life, 2021). This indicates how optimistic young people are when predicting their likely exam grades. The generosity in grading and the different ways in which assessment was conducted for National Qualifications in pandemic years raises questions about how the grades can be interpreted and whether this can be compared with normal years. SQA statistical releases stated publicly that the grades were not comparable in terms of standard of the outcomes with previous years. There remains, though, the question of whether the grades were at a different level, but their meaning could be interpreted in the same way, or if they represented something different altogether.

A **cohort referencing** approach to defining standards simply gives a certain proportion of candidates the grades each year (Wiliam, 1996). For example, 10% of candidates might be given a grade A each year, no matter the levels of performances observed on the assessments. With the rise in outcomes over the pandemic, it is clear that a cohort referenced approach was not applied in awarding SQA qualifications in 2020 and 2021.

Some might not want to grapple with an abstract meaning of grades and instead use a **due process** approach to defining standards (Cizek, 1993). In such a definition, the grades are simply the outcomes awarded following the pre-specified processes for assessment administration. As we saw from the furore surrounding Results Day in 2020, it is unlikely that such a definition would gain public confidence. First, stakeholders questioned the authority of the processes and second, they questioned the warrant upon which the grading has been issued. A due process definition of standards does not describe how people understand and accept grading outcomes for national qualifications.

Others might be attracted to the notion that grades tell us something about the underlying ability of the candidates, such as ability in science or even intelligence. This is a **construct-referenced** approach, since the grades indicate the level of ability on the construct being assessed (Wiliam, 1996). With grade increases during the pandemic, at the same time as learning loss was being documented across a range of studies (EEF, 2021), a construct-referenced approach cannot be the correct way to interpret grading outcomes. If young people had less opportunity to learn, their underlying skills would not have been so developed and the outcomes should have been lower.

SQA's policy on standards indicates that graded national qualifications (National 5, Highers and Advanced Highers) have an **attainment-referencing** definition (Newton, 2011; Baird et al., 2000). This approach takes into account students' attainments as indicated by their performances in the assessments and the statistics used to evaluate the assessment outcomes. This is similar to the definition used for GCSE and A-level examinations in England, Wales, and Northern Ireland (Taylor and Opposs, 2018) and for the Leaving Certificate examinations in Ireland (McManus, 2018). An attainment-referencing definition leads to greater stability in outcomes at the cohort level than in some of the other definitions of standards, meaning that the education system can operate with expectations regarding the grades that will be issued. A disadvantage with this approach in comparison with a criterion-referenced approach to defining standards is that there is no single artefact that indicates the standard. Attainment-referencing takes into account a variety of sources of qualitative and quantitative information. Students' performances might look better in a given year, but this could have been caused by an easier question paper. Grade boundaries can then be adjusted to account for a change in assessment demand under the attainment-referencing definition. This ensures that the marks associated with corresponding grade boundaries represent equivalent levels of attainment year-on-year. Arguably, the SQA statistical moderation mechanism was an attempt to use attainment referencing with teacher estimates and statistical information about the performance of students in each school or college in previous years. The SQA statistical moderation was abandoned in 2020 because it was seen to be unfair. This does not mean that attainment referencing would be seen to be unfair in normal times, only that this particular mechanism, the use of the statistical moderation, was not accepted. The results in 2020 and 2021 were driven by teacher estimates and were not attainment referenced because cohort-level statistics were not brought to bear on producing the grading distributions.

Yet another approach to defining standards is **criterion referencing** (Popham and Husek, 1969). Under this definition, people deserve the grades awarded if they meet the pre-determined criteria, which are written in text, often in the form of grading criteria. The conditions of assessment varied under the pandemic, but it would be feasible to say that teachers estimated whether students met the criteria for grades. Grade inflation could then be explained by optimism or by a focus on students' best performances when making grading judgments. This definition is compatible with grade inflation because no statistical methods are used, and the outcomes are simply an aggregate of the assessors' judgments. A feature of the criterion referencing method of standard setting is instability in the outcomes for national qualifications (Baird, 2007).

Across all three methods of data collection in the current research, stakeholders generally prioritised a criterion-referenced definition of standards. The following quotation from a college student illustrates how this was discussed in one focus group,

> *The standards are obviously the guidelines that they are setting out that they want, because obviously within the different levels of education there will be different standards for different levels. Obviously, the higher the level of education, ... higher the standards will need to be.*

(College student)

However, there were contradictory views and a substantial proportion (40%) of questionnaire respondents considered that grades represented an underlying ability, not just student performances on the assessment, which is closer to a construct-referenced view of standards. The use of statistics in standard setting was acknowledged by participants, but the questionnaire showed that a low proportion of respondents considered that keeping the statistical outcomes similar across years was a priority. This suggests that a reliance on statistical methods in standard setting, such as a cohort-referencing approach, would not be supported as the method of choice by most participants. Although the attainment-referenced approach currently used in Scotland and elsewhere relies on both statistical and other sources of information, the approach may not be fully understood by participants. However, the implications for the value of qualification grades if outcomes were not deemed to be consistent across years may not have been foremost in the minds of stakeholders when answering these questions during a pandemic. Further, the need for both statistical information and expert judgment of examinees' performances in a standard setting process may not be clear to those who are not national assessment specialists. Using both sources of information is the international norm for national qualification standard setting procedures (Cizek and Bunch, 2007).

There was a general view in the PASS project questionnaire data that the National Qualifications had different standards during the pandemic. Although the outcomes had been higher, there was a nuanced view about the extent to which this meant that they were easier. Stakeholders recognised the challenges faced by students and how this affected their learning opportunities. Only a quarter of employers thought that the standards were lower in 2021 compared with a normal year, though many employers did not feel that they were able to judge this matter. Grade inflation was therefore not a main concern during this period, though some college representatives in the focus groups expressed a view that teacher assessment in schools was responsible for grade inflation. College lecturers considered that their more extensive experiences of moderation put them in a better position to avoid grade inflation during the pandemic.

The generous approach taken to grading in the pandemic in the UK can be seen as retributive fairness (see the section on fairness below), as a reaction to the special circumstances of these years. Further aspects of retributive fairness can be seen in the way assessment was undertaken in those years, with teachers being able to focus upon only those aspects of the course that were taught, advance notice of questions and other techniques being used to reflect the fact that less time was spent in school. Ofqual's (2020c) instructions to schools in England were that the outcomes should "reflect the grades students would have been most likely to achieve if teaching and learning had continued and they had taken their exams as normal." Thus, the grades in England were a prediction, based upon the evidence available to teachers. All these adaptations challenge the traditional definitions of standards, though a 'due process' was in place and criterion referencing was used in each case. The criteria, though, were adapted to suit the circumstances of each assessment and, since the evidence base was different across schools in many cases, the criteria would have been manifested differently.

Trust and communication

Public confidence in national assessments is of the utmost importance if qualifications are to have value. The Rapid Review of National Qualifications Experience 2020 concluded that stakeholders perceived a lack of transparency in how the processes were communicated, which affected trust in the system when results were released (Priestley et al., 2020). Of course, examination crises have occurred previously, internationally and in Scotland (see, for example, Baird and Coxell, 2009), leading to considerations of how public confidence can best be upheld. Historically, examination boards were not transparent in their methods. Instead, there was an assumed authority. Many societies have changed, however, and there is less deference to authority generally and institutions are now questioned and must justify their procedures and decisions. Indeed, there have been calls for greater public understanding of assessment such that the public is better able to interpret and engage with debates related to assessment (e.g., Newton, 2005a).

Media narratives position assessment as something which ought to be error-free (Newton, 2005b) and there have been calls for better management of the media by examination bodies to counteract this (Warmington and Murphy, 2007). Ultimately all assessment includes error to greater or lesser degrees. In 2008, the examination regulator in England set out to address the lack of knowledge concerning assessment reliability in qualifications through a programme of work called the *Reliability Programme*. Newsworthy stories arising from this programme were not always those that an examination body would wish to promote, including sensationalising the findings or positioning them as whistleblowing and presenting the authorities as trivialising or evading the issues (Newton, 2013). Lessons have been learned and communication strategies are now commonplace amongst examination boards. However, public confidence is not simply a matter of good PR, nor is it solely related to measurement error or reliability issues. Transparency and trust are not necessarily complementary; openness may actively undermine trust since it can draw attention to problems - apparent or real (O'Neill, 2002).

Research on stakeholders' views of trust in public examinations in England has shown that the main concerns went beyond marking error, and included examination standards, the syllabus, communication with the examination board, fitness for purpose of the assessment and the mode of

assessment, such as modular structures (Simpson and Baird, 2013). Further, stakeholders' concerns on assessment reliability specifically were quite different from those of industry insiders who manage the examinations and seek to promote public understanding and confidence (Chamberlain, 2013). Stakeholders did not engage well with distinctions between random and human error and did not find the notion of measurement error relevant or useful. In another survey of teachers, students and employers in England, most respondents indicated that any level of error was unacceptable (He et al., 2011).

Despite the media narratives surrounding national examinations, research shows that public confidence in national examinations tended to be high. An annual survey of stakeholder views in England consistently found that approximately three-quarters of respondents trusted GCSE examinations and the figure for A-levels was even higher at around 80%. In 2020, trust in these two qualifications fell to 27% and 29% respectively (YouGov, 2021). Similarly, in Scotland in 2019, 9 out of 10 respondents in SQA's stakeholder tracking research programme considered that SQA could be trusted and 93% considered that the Higher qualifications had high credibility. These are extraordinarily high levels of trust. System shocks can have a substantial impact upon trust, though, and it can take several years to regain that public trust. Stakeholders' views on what is important for trust are multifarious and includes having the right expertise in examination boards, the examination system being a national one, use of statistical procedures to make sure the results are accurate, and the examinations assessing the right things at the right level (He et al., 2011). In 2020, there was not only an erosion of trust in the qualifications in Scotland, but trust between young people and their teachers was undermined because the results were based upon teacher judgments (Priestley et al., 2020). A pupil in one of our focus groups said,

> *But the system as a whole, I think there is quite a lot of trust and confidence that has been lost, which might be quite damaging for years ahead. […] I think it is very difficult now to try build that back.*
>
> (School pupil)

Nonetheless, our questionnaire findings showed that only one third of respondents in Scotland did not trust the Scottish assessment results in 2021. At the time of the research (November 2021), there was a reported decline in trust in SQA amongst most stakeholders. Questionnaire respondents got their information about National Qualifications from a variety of sources, including the schools and colleges, SQA and personal contacts, as well as from the news and other media. In contrast to other stakeholder groups, the employer survey respondents' trust in SQA was largely unchanged (71 of 103) and an exceedingly small number reported a decline in trust (9 of 103).

Teacher assessment

Teacher assessments took particular forms during the pandemic. In normal years there is much more time for planning, administration and conduct of teacher assessments, which better underpins their quality. A number of models in which teacher assessments are integrated in final qualification results have been documented and evaluated (Wilmut, 2005). Nevertheless, teacher assessment has lower reliability than examinations, since processes surrounding external examinations are generally more controlled and standardised (Johnson, 2013). Teacher assessments tend to produce higher grades than external examinations (Wikström and Wikström, 2005). Various approaches to improving the

standardisation of teacher assessments have been used, including statistical moderation (Wilmut and Tuson, 2005). Different moderation techniques were applied to the ACM results during the pandemic in 2020 and 2021, as outlined above. The statistical moderation technique applied in 2020 was not acceptable to stakeholders.

Although teacher professionalism is respected in Scotland, there were concerns about potential bias in teacher assessment in 2020 (Priestley et al., 2020). Research evidence on this can be drawn from experiments designed to investigate bias, from comparisons between teacher assessments and examination results and from comparisons on teachers' estimates or predictions of students' results with their actual results. A review of the international literature on bias in teacher assessment drew the following general conclusions from the available research on comparisons between examination results and teacher assessments or predictions (Lee and Walter, 2020):

- Accuracy of teacher predictions varied to a small extent by age and gender, but this was inconsistent across subjects
- Teacher predictions varied by subject, but the effects were inconsistent
- Accuracy of prediction was related to the kind of school that pupils attended
- Predictions for some ethnic minority groups were overly optimistic
- There was some over-prediction for disadvantaged groups
- Amongst high-attainers, there was less over-prediction for disadvantaged groups

Thus, young people's concerns about the use of teacher judgment in the production of their qualification result potentially being biased are to some extent justified by the existing research literature. However, examination results differ by subject, school type, ethnic minority, age, gender, disadvantage and, of course, by prior attainment. So, the question is not whether there are patterns in teacher assessments relating to these variables, but whether they are exacerbated in teacher assessments. In England, an equality analysis showed that the teacher-produced Centre Assessed Grades had not altered the attainment gaps for the most part (Stratton, Zanina and Noden, 2021). Interestingly, the subjects which had the highest increases in grades in England in 2020 were those in which teacher assessment was usually part of the final grade. This may have been due to teachers' estimates being most influenced by assessments of the part of the syllabus that was usually assessed by teachers, and which was typically graded most generously. Indeed, the finding that attainment gaps were not affected by the 2020 procedures in England pertained when the statistical moderation was applied to the results and when teacher assessments were used (Lee, Stringer and Zanini, 2020).

Teachers are highly trusted in Scotland (Gray and Weir, 2014). Conflicts of interest due to accountability systems, such as league tables, have made teacher assessment more problematic in other settings (Ofqual, 2013). Secondary school teachers in England who responded to a survey indicated that they had observed the following behaviours amongst colleagues: giving hints to pupils during controlled assessments, giving students the benefit of the doubt when assessing coursework and giving students specific wording to use in sections of their coursework (Meadows and Black, 2018). Whilst league tables are not an official feature of the education system in Scotland, schools' and teachers' reputations are associated with assessment outcomes and teachers are likely to feel pressure to produce good outcomes for their pupils from a range of sources, including parents. Moreover, changes to the assessments during the pandemic radically altered the teacher-pupil relationship in a manner that could not be carefully managed and planned over time. Fairness of teacher assessments is often contrasted with fairness of examinations in the literature and in the

discussions regarding assessment policy under COVID-19. Interestingly, 45 of the 103 employers in the PASS project survey reported that examinations were, in general, the best way to assess learners.

There was overall support for the use of teacher assessment during the pandemic in our data. Employers, too, considered that the use of teacher grading was the best policy under the circumstances. The questionnaire data indicated that most respondents felt that teachers and lecturers were well-prepared to make assessment decisions in 2021. Opinion was divided, though, as to whether teacher assessment was fair. There was a view that bias arising from teacher-pupil relationships or parental pressure could have affected grades. Where teacher assessment is used, there is a need for a well-developed moderation system to address potential bias. The urgency associated with the pandemic necessarily meant that the systems in place were devised and implemented speedily. They built upon existing quality assurance approaches within schools, which required that decisions for internal assessments were moderated within schools and colleges. Since 2016, there was a reduction in internally assessed units from National Qualifications at National 5, Higher and Advanced Higher. Workload for teachers arising from teacher assessment is another issue that will need to be considered if future reforms build upon the teacher assessment systems put in place during the pandemic. A journalist interviewed in the focus group study made the following observations,

> *Gordon Stobart made [the point] in his report ... we had tried to have more teacher assessment in our qualifications when they were first introduced ... the ones that were designed specifically to chime with Curriculum for Excellence. And the SQA ... defended itself ... saying that - initially there was not quite so much focus on... a high stakes exam because you had the unit assessments, and ... teachers played a bigger role in assessing pupils ... Well, we know that all of that then fell apart, because teachers complained about workload, but so did pupils ... So... learn the lesson of that as we move forward to look at how our qualifications could be in the future*

<div align="right">(Journalist)</div>

Fairness

Within the research literature, fairness came to be defined in a way that would not be recognised by stakeholders (Nisbet and Shaw, 2019). It was driven by legal cases in the US and statistical approaches (differential item functioning analysis) to identifying group differences in performance on individual questions have been used a lot in the exploration of possible bias. In such approaches, the concerns of the individual were overlooked (Shaw and Nisbet, 2021). Management of an assessment system is conducted at the aggregate level. Considerations of relational fairness are typically in relation to groups, because tackling relational fairness is a practical prospect. Comparisons of individuals without categorisation would become an unmanageable task with a small group of pupils, let alone a national system.

Through a broader, philosophical consideration, four meanings of the term fairness were discerned by Nisbet and Shaw (2019, 2020). Their first two forms of fairness are categorised as 'technical' ways of viewing the notion. In *formal* fairness, the concern is simply that a rule has been applied accurately, as intended. There is no questioning, as such, of the rule itself. A grade would be fair if the rules for grading had been applied as outlined in the regulations. In *implied contractual* fairness,

something is fair if it meets the legitimate expectations of those affected. Such expectations may arise from regular practices or from express promises or policy which people can reasonably expect to continue. The second two definitions of fairness are termed 'ethical' ways of viewing fairness. In *retributive* fairness, something is fair if the consequences are appropriate, given what has gone before. As such, an assessment grade would be fair in an attributive sense if it rewarded (or punished) the educational performances leading up to it. Performances or hard work might be considered when viewing an assessment as retributively fair, for example. In *relational* fairness, something is fair if relevantly like cases are treated alike. This appeal to fairness arises most often in situations of competition for resources. For example, access to prestigious university places might give rise to arguments about the relational fairness of assessments.

The Victorian Curriculum and Assessment Authority in Australia adopted a distinctive approach to learning loss during the pandemic.[1] Teachers were required to estimate the learning loss suffered by young people and on this basis, adjustments were made to their grading outcomes. Thus, the grading reflected what teachers estimated that students would have attained through this mechanism. This can be seen as a retributive sense of grading fairness, and it was widely accepted in Victoria. How accepted this approach would have been in other countries is an open, and highly political, question.

In 2020, much of the public discussion regarding fairness in Scotland centred upon the lack of transparency of the assessment and moderation process (Priestley et al., 2020). Pupils felt that they did not have the opportunity to demonstrate their knowledge and skills in the traditional manner and were assessed on class work which had been carried out without the usual understanding of the stakes associated with them and enough prior notice to prepare for them. Due to the lockdown, assessments were made based on work already completed. The common nightmare of having to take an exam without preparation was turned on its head – instead, young people found that they had been unwittingly examined. As Shaw and Nisbet (2021, p13) noted, "Rarely have exams been so praised as in their absence." In the TeenCOVID-19Life (2021) survey of 671 young people who received their results in Scotland in August 2020, 69% thought that replacing the SQA statistically moderated grades with teacher estimates was fair and 86% were happy with their results. In 2021 there was less of a public response to the results in terms of fairness, as teacher assessments were used and there was greater transparency and planning associated with the system. In the PASS project data, only one third of questionnaire respondents indicated that they felt the 2021 results were unfair and only ten of the 103 employers thought this.

Writing about Scotland, England and Ireland, Shaw and Nisbet (2021) argued that relational fairness came to the fore during the pandemic, since many of the criticisms stemmed from the notion that students felt they were being treated less well than others like themselves. Social media no doubt contributed to these feelings. Relational fairness concerns regarding social disadvantage in the PASS project data are illustrated in the following quotation from a parent in one of the focus group discussions:

> *I speak really strongly about Scottish Borders because it is where I am from, I am very proud about it, but I felt they were poorly treated because of the type of area that we live in.*

Which comparison is central is important here since many different axes of comparison can be made; such as ethnicity, gender, geographical, socioeconomic status. Further, Shaw and Nisbet (2021) noted that learning loss discussions gave more impetus to a retributive sense of fairness, even though equality gaps in access to education, as well as in outcomes, prevail in normal times. The equality impact assessment of the 2020 results conducted by the Scottish Government did not find any direct or indirect discrimination based on the use of teacher estimates, though there were clearly rises in outcomes (Scottish Government, 2020b). Shaw and Nisbet (2021) argued that standards over time were less of a focus during the pandemic due to the special circumstances and question whether this will be a new normal, or if concerns related to maintenance of standards across years will resurface.

Relational fairness was also raised in the focus group data from the PASS project, as standardisation and consistency were seen by many participants as key features of a fair system. Consistent assessment criteria, moderation and verification processes were mentioned in the discussions. Additionally, having the same assessments and specifically an examination format was raised. Standardisation and inclusion agendas are in tension here though, as, if diversity is to be addressed, some variation in provision is important. Some focus group participants considered that National Qualifications had a role in perpetuating inequalities, beyond simply reflecting wider inequity in the education system and in society more broadly.

The future of assessment in Scotland

As outlined in the introduction to this report, this research was conducted whilst several reviews of the National Qualifications system in Scotland were being undertaken. National debates are therefore likely to have influenced our findings. Eighty percent of respondents to the stakeholder questionnaire indicated that reform was needed, though employers in our survey were not as supportive of reform. Any reform of assessment must start with a consideration of the purpose of assessment, a question we put to survey participants. The most important purpose for National Qualifications selected was as a source of information for the learner rather than information for employers, selection for further or higher education, or the evaluation of schools or the education system.

There was a call for greater choice, both for schools and individual learners in form of assessment from 60% of questionnaire respondents. Focus group discussions, conducted earlier in 2021, indicated a desire to reduce the reliance on end of course examinations in assessments. In fact, except for Latin and mathematics, National Qualifications in Scotland do not rely on a single end of course examination, though many stakeholders seemed to think that they do. A national system which integrates examinations and teacher assessment was supported in the stakeholder questionnaire data.

The vision of assessment that came through the data was learner-focused, with choice, flexibility, and the capacity to adapt to the needs of learners with different skills and in a range of contexts. In terms of process, there was a desire for better representation of teachers' and learners' voices in future reforms. Improved communication was raised as a significant issue for public confidence that

should be a legacy of the pandemic. Although examiners in Scotland tend to be teachers and teachers are often involved in assessment reform, there was a call for more involvement of teachers. Nonetheless, one of the practitioners who responded to the stakeholder questionnaire encapsulated a theme that was mentioned several times:

> *Having more teachers involved at design stage and present on high level boards at SQA would massively support a more coherent approach to education and assessment across Scotland*

(Teacher/Lecturer)

The vision of assessment from stakeholders that came through the data was learner-focused, with choice, flexibility, and the capacity to adapt to the needs of learners with different skills and in a range of contexts.

References

Baird, J. (2007) Alternative conceptions of examination standards. In Newton, P., Baird, J., Goldstein, H., Patrick, H. and Tymms, P. (Editors) (2007) *Techniques for monitoring the comparability of examination standards*. QCA book publication. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/487054/2007-comparability-exam-standards-f-chapter4.pdf

Baird, J. and Coxell, A. (2009) Policy, Latent Error and Systemic Examination Failures. *CADMO*, XVII, 2, 105-122.

Baird, J., Cresswell, M.J. and Newton, P.E. (2000). Would the real gold standard please step forward? *Research Papers in Education*, 15, 2, 213 – 229.

Chamberlain, S. (2013) Qualification users' perceptions and experiences of assessment reliability, *Research Papers in Education*, 28:1, 118-133, DOI: 10.1080/02671522.2012.754231

Cizek, G.J. (1993) Reconsidering standards and criteria. *Journal of Educational Measurement*, 30 (2), 93–106.

Cizek, G.J. and Bunch, M.B. (2007) *Standard setting: a guide to establishing and evaluating performance standards on tests.* Sage Publications Ltd.

Education Scotland (2021) *Equity Audit. Deepening the understanding of the impact COVID-19-19 and school building closures had on children from socio-economically disadvantaged backgrounds and setting clear areas of focus for accelerating recover*. https://education.gov.scot/media/2ygfjxhd/equityaudit.pdf

Education Endowment Fund (EEF) (2021) *Best evidence on impact of COVID-19-19 on pupil attainment*. https://educationendowmentfoundation.org.uk/guidance-for-teachers/COVID-19-19-resources/best-evidence-on-impact-of-COVID-19-19-on-pupil-attainment

Gray, D. and Weir, D. (2014) Retaining public and political trust: teacher education in Scotland, *Journal of Education for Teaching*, 40:5, 569-587, DOI: 10.1080/02607476.2014.956541

He, Q., Boyle, A. and Opposs, D. (2011) Public perceptions of reliability in examination results in England, *Evaluation and Research in Education*, 24:4, 255-283, DOI: 10.1080/09500790.2011.629722

Johnson, S. (2013) On the reliability of high-stakes teacher assessment, *Research Papers in Education*, 28:1, 91-105, DOI: 10.1080/02671522.2012.754229

Lee, M.W. and Walter, M. (2020) *Equality impact assessment: literature review*. Ofqual Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879605/Equality_impact_assessment_literature_review_15_April_2020.pdf

Lee, M.W., Stringer, N. and Zanini, N. (2020) Student-level equalities analyses for GCSE and A level. Ofqual Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938869/6713_Student-level_equalities_analyses_for_GCSE_and_A_level.pdf

Major, L.E., Eyles, A. and Machin, S. (2021) *Learning loss since lockdown: variation across the home nations.* Centre for Economic Performance. COVID-19-19 Analysis Series. No.023. July. https://cep.lse.ac.uk/pubs/download/cepCOVID-19-19-023.pdf

McManus, H. (2018) Chapter 9 – Standard setting in Ireland: The Leaving Certificate. In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (Editors) *Examination standards: How measures and meanings differ around the world.* UCL IOE Press. 157 – 177.

Meadows, M. and Black, B. (2018) Teachers' experience of and attitudes toward activities to maximise qualification results in England, *Oxford Review of Education*,44:5, 563-580, DOI: 10.1080/03054985.2018.1500355

The Muir Report (2022) *Putting Learners at the Centre: Towards a Future Vision for Scottish Education* Scottish Government available at https://www.gov.scot/publications/putting-learners-centre-towards-future-vision-scottish-education/ (Accessed 23/5/22)

Newton, P.E. (2005a) The public understanding of measurement inaccuracy. *British Educational Research Journal*, 31, 4, 419 – 442. https://doi.org/10.1080/01411920500148648

Newton, P.E. (2005b) Threats to the professional understanding of assessment error, *Journal of Education Policy*, 20:4, 457-483, DOI: 10.1080/02680930500132288

Newton, P.E. (2011) 'A level pass rates and the enduring myth of norm- referencing'. *Research Matters*, Special Issue 2, 20–6.

Newton, P.E. (2013) Ofqual's Reliability Programme: a case study exploring the potential to improve public understanding and confidence, *Oxford Review of Education*, 39:1, 93-113, DOI: 10.1080/03054985.2012.760285

Nisbet, I. and Shaw, S.D. (2019) Fair assessment viewed through the lenses of measurement theory, Assessment in Education: Principles, Policy & Practice, 26:5, 612-629, DOI: 10.1080/0969594X.2019.1586643

Nisbet, I. and Shaw, S.D. (2020) *Is Assessment Fair?* London. Sage Publications Ltd.

OECD (2021a) *The state of School Education: One year into the COVID-19 pandemic*. Paris: Organisation for Economic Co-operation and Development. https://doi.org/10.1787/201dde84-en

OECD (2021b) *Scotland's Curriculum for Excellence: Into the Future*, Implementing Education Policies, OECD Publishing, Paris, available at https://doi.org/10.1787/bf624417-en. (Accessed 23/5/22)

Ofqual (2013) *Review of controlled assessments in GCSEs*. (Coventry, Ofqual). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/377903/2013-06-11-review-of-controlled-assessment-in-GCSEs.pdf

Ofqual (2020) *Fairness in Awarding*. The Ofqual Blog. https://ofqual.blog.gov.uk/2020/08/06/fairness-in-awarding/

O'Neill, O. (2002) *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press, Cambridge.

Opposs, D. (2020) *The impact of the coronavirus outbreak on exams around the world*. The Ofqual Blog. 22 May. https://ofqual.blog.gov.uk/2020/05/22/the-impact-of-the-coronavirus-outbreak-on-exams-around-the-world/

Popham, W. J., and Husek, T. R. (1969) Implications Of Criterion-Referenced Measurement. *Journal of Educational Measurement*, *6*(1), 1-9.

Priestley, M., Shapira, M., Priestley, A., Ritchie, M. and Barnett, C. (2020) *Rapid Review of National Qualifications Experience 2020.* Final Report. September. Scottish Government. https://www.gov.scot/publications/rapid-review-national-qualifications-experience-2020/

Scottish Government (2020a) *Deputy First Minister - SQA 2020 results* Statement given by Education Secretary John Swinney to the Scottish Parliament on 11 August 2020 https://www.gov.scot/publications/deputy-first-minister-sqa-2020-results/

Scottish Government (2020b) *Equality Impact Assessment: 2020 Results – Teacher and Lecturer Estimates.* December. https://www.gov.scot/binaries/content/documents/govscot/publications/impact-assessment/2020/12/coronavirus-COVID-19-19-equality-impact-assessment-2020-results-using-teacher-lecturer-estimates/documents/equality-impact-assessment-2020-results-teacher-lecturer-estimates/equality-impact-assessment-2020-results-teacher-lecturer-estimates/govscot%3Adocument/equality-impact-assessment-2020-results-teacher-lecturer-estimates.pdf

Shaw, S.D. and Nisbet, I. (2021) Attitudes to fair assessment in the light of COVID-19-19. *Research Matters: A Cambridge Assessment publication*, 31, 6 – 21. https://www.cambridgeassessment.org.uk/Images/https-cambridgeassessment.org.uk-images-research-matters-31-attitudes-to-fairassessment-in-the-light-of-COVID-19-19.pdf.pdf

Simpson, L and Baird, J. (2013) Perceptions of trust in public examinations, Oxford Review of Education, 39:1, 17-35, DOI: 10.1080/03054985.2012.760264

SQA (2020a) *2020 Alternative Certification Model: Equality Impact Assessment* SQA https://www.sqa.org.uk/sqa/files_ccc/2020-sqa-alternative-certification-model-equality-impact-assessment.pdf

SQA (2020b) *National Qualifications 2020 Awarding — Methodology Report* SQA
https://www.sqa.org.uk/sqa/files_ccc/SQAAwardingMethodology2020Report.pdf

SQA (2020c) *Outcome of the technical consultation on the proposed modifications to National 5, Higher and Advanced Higher course assessments for session 2020–21: Analysis of consultation responses* SQA https://www.sqa.org.uk/files_ccc/outcomes-national-consultation-2021.pdf

SQA (2021a) *National Qualifications 2021 Alternative Certification Model (ACM): Methodology Report* SQA https://www.sqa.org.uk/sqa/files_ccc/SQAAwardingMethodology2021Report.pdf

SQA (2021b) *2021 National Qualifications Appeal Outcomes* SQA
https://www.sqa.org.uk/sqa/files_ccc/sqachiefexaminingofficer2021reportappeals.pdf

Stobart, G. (2021), "Upper-secondary education student assessment in Scotland: A comparative perspective", *OECD Education Working Papers*, No. 253, OECD Publishing, Paris, available at https://doi.org/10.1787/d8785ddf-en (Accessed 23/5/22)

Stratton, T., Zanin, N. and Noden, P. (2021). An evaluation of centre assessment grades from summer 2020. Ofqual. https://www.gov.uk/government/publications/evaluation-of-centre-assessment-grades-and-grading-gaps-in-summer-2020/an-evaluation-of-centre-assessment-grades-from-summer-2020

Taylor, R. and Opposs, D. (2018) Chapter 6 – Standard setting in England: A levels. In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (Editors) *Examination standards: How measures and meanings differ around the world.* UCL IOE Press. 100 – 113.

TeenCOVID-19Life (2021) TeenCOVID-19Life Survey 2 – Schooling Mini Report. Cancelled SQA exams – impact on young people. February. Schools Health and Wellbeing Improvement Research Network. Generation Scotland. https://www.ed.ac.uk/generation-scotland/what-have-we-found/latest-results/teenCOVID-19life2-results

UNESCO (2020) COVID-19-19. A glance of national coping strategies on high-stakes examinations and assessments. Working Document. 11 April.
https://en.unesco.org/sites/default/files/unesco_review_of_high-stakes_exams_and_assessments_during_COVID-19-19_en.pdf

Urhahne, D. and Wijnia, L. (2021) A review on the accuracy of teacher judgments, *Educational Research Review*, 32, Article 100374. https://doi.org/10.1016/j.edurev.2020.100374.

Warmington, P. and Murphy, R. (2007) 'Read all about it!' UK News Media Coverage of A-Level Results. *Policy Futures in Education*, 5, 1, 70 – 83. https://doi.org/10.2304/pfie.2007.5.1.70

Wikström, C. and Wikström, M. (2005) Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. *Economics of Education*, 24, 3, 309 – 322. https://doi.org/10.1016/j.econedurev.2004.04.010

Wiliam, D. (1996) 'Standards in examinations: A matter of trust?'. *Curriculum Journal*, 7 (3), 293–306.

Wilmut, J. (2005) *Experiences of Summative Teacher Assessment in the UK*, London: Qualifications and Curriculum Authority. [Google Scholar]

Wilmut, J., and Tuson, J. (2005) *Statistical moderation of teacher assessments: A report to the Qualifications and Curriculum Authority*. London: QCA. [Google Scholar]

Wolf, A., and Jenkins, A. (2005) Employers' selection decisions: the role of qualifications and tests. In *What's the Good of Education? The Economics of Education in the UK* (pp. 147 - 168). Princeton University Press.

YouGov (2021) Perceptions of AS and A levels, GCSEs, and Applied General qualifications in England – Wave 19. An Ofqual commissioned report. April. Ofqual/21/6777/1. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/980782/Perceptions_Survey_Wave_19_-_Report.pdf