

Perceptions of Assessment Standards in Scotland: Focus Groups with Stakeholders

Report 2 of 4

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

September 2022



Contents

1	EXECUTIVE SUMMARY	1
1.1	ABOUT THE PROJECT	1
1.2	THIS REPORT	1
1.3	DATA	1
1.4	FINDINGS	1
1.5	ACKNOWLEDGEMENTS.....	2
2	PERCEPTIONS OF ASSESSMENT STANDARDS IN SCOTLAND PROJECT	3
2.1	UNDERSTANDING STAKEHOLDER VIEWS OF QUALIFICATIONS IN SCOTLAND.....	3
2.2	STAKEHOLDERS' PERCEPTIONS OF ASSESSMENT STANDARDS IN SCOTLAND	3
2.3	APPROACH TO THE RESEARCH.....	3
3	FOCUS GROUP METHODOLOGY.....	3
3.1	DATA COLLECTION.....	3
3.2	PARTICIPANT RECRUITMENT	4
3.3	CONDUCT OF THE FOCUS GROUPS.....	5
3.4	ANALYTICAL STRATEGY	6
3.5	ETHICS	7
4	FOCUS GROUP FINDINGS	8
4.1	PARTICIPANTS.....	8
4.2	WHAT INFLUENCED THE REACTIONS OF DIFFERENT COMMUNITIES TO THE SQA QUALIFICATIONS PROCESSES AND RESULTS IN 2020?.....	9
4.2.1	How and why did communities react to the qualifications process	9
4.2.2	The role of the media.....	10
4.2.3	Reactions to the final results.....	11
4.2.4	What does transparency mean in this context?.....	13
4.3	WHAT FACTORS INFLUENCE DIFFERENT COMMUNITIES' PERCEPTIONS OF FAIRNESS FOR ALL LEARNERS IN QUALIFICATIONS?	13
4.4	HOW DO PEOPLE IN DIFFERENT COMMUNITIES UNDERSTAND THE TERM STANDARDS IN THE CONTEXT OF QUALIFICATIONS?	16
4.5	WHAT DO PEOPLE IN DIFFERENT COMMUNITIES BELIEVE CAN BE LEARNT FROM EXPERIENCES IN 2020 AND 2021 IN RELATION TO FAIRNESS, STANDARDS AND PUBLIC PERCEPTIONS?.....	20
4.5.1	What can be learnt to ensure future qualifications are fair and that standards are dependable?.....	20
4.5.2	What actions are required to inform public perceptions of future qualifications.....	22
4.6	CONCLUSIONS	22
5	REFERENCES	25

List of Tables

Table 1	Focus group participants.....	8
Table 2	Coding data relating to the reactions of different communities to the 2020 qualification process	9
Table 3	Definitions of fairness coded, from Nisbet and Shaw 2019, 2020	13
Table 4	Coding data relating to the factors influencing perceptions of fairness	14
Table 5	Coding data on factors which make an assessment fair.....	14
Table 6	Definitions of standards coded	17
Table 7	Coding data relating to understanding of standards.....	17
Table 8	Coding data relating to what can be learnt from experiences in 2020 and 2021	20

1 Executive Summary

1.1 About the project

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 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 policy-making and communication with stakeholders. The research was conducted by teams from the University of Glasgow, the University of Oxford and SQA.

1.2 This report

Focus group discussions with a broad range of 82 stakeholders were conducted first. These data form the basis for the current report. Views of school pupils, college students, teachers, lecturers, parents or carers, senior management team members in schools and colleges, employers, university admissions officers and journalists were sought.

1.3 Data

The focus group discussions took place between April and June 2021. Nineteen focus groups were conducted with participants from 27 schools, 14 colleges, three higher education institutions, three small and medium businesses and, due to drop-out, only one media organisation. The discussions were recorded and transcribed. They were then analysed qualitatively to produce rapid reports, with draft findings. These rapid reports informed both SQA's imminent policy discussions and coding of the data in NVivo software. Findings from the focus groups informed the design of the subsequent questionnaire and telephone survey.

1.4 Findings

Although most participants recognised that the shift to teacher grades represented the best solution in a difficult situation, a lot of negative reactions to the way in which the 2020 assessment policy had been handled were expressed. This was less so for 2021, though significant issues were still raised. There was anger, frustration and disappointment relating to both 2020 and 2021 arrangements. Lack of timely communication was cited as highly problematical. Participants recognised that the media had a role in communication and did not always agree with the narratives presented. Much of the discussion related to the use of the statistical moderation that had been initially applied to results in summer 2020, which came to be known as the 'algorithm'. The 'postcode lottery', referring to the use of a school or college's prior attainment data in the 'algorithm', was seen as unfair in media narratives and a number of focus group participants also referred to these terms.

Following the Government policy change after Results Day, results were issued on the basis of teacher grades. There was a lot of support for the use of teacher grades and for the final results arising from them in 2020. Some participants from the college sector considered that the use of teacher assessments in the final

results of 2020 favoured the school sector and that school teachers were more responsible for grade inflation than the college sector, which was accustomed to conducting moderation.

Standardisation and consistency were mentioned by many participants as key factors underpinning fairness. By these terms, they had in mind having consistent assessment criteria, moderation and verification processes to ensure that these were applied in the same way across institutions and the circumstances under which learners sat assessments. As such, examinations were mentioned favourably in terms of the standardisation that they offered. There was a recognition that national assessments cannot wholly address inequalities, though some participants considered they may have a role in constructing and perpetuating some of them. On the whole, there is a high emphasis on ethical fairness from all participants, although surprisingly learners did focus on fairness in a more technical sense – with reference to accuracy or whether it met the legitimate expectations of stakeholders such as university admissions officers.

The overwhelming majority of the discussion about the meaning of standards related to a criterion-referenced definition. In this approach, performance requirements are outlined in advance of the assessment and performances are judged by suitably qualified examiners. Statistics are not used in a criterion-referenced system to check whether the results are plausible or accurate. Some stakeholders discussed the need for a system-level check on outcomes and an expectation that statistics were likely to remain stable at a national level, year on year. This is a cohort-referenced definition of standards. Notably absent was any discussion of an attainment-referenced approach, in which professional judgements of performances and national statistics are used jointly to set standards. However, it is entirely possible that the need for both elements is not clear to those who are not national assessment specialists.

Many suggestions were made regarding what could be learned from the pandemic in relation to assessment matters. There was a lot of discussion about the opportunity for reform of qualifications and whether the system was overly reliant on end of course examinations. Given the absence of examinations during the pandemic, stakeholders were keen to see them retained in a future system to at least some extent. Issues relating to appeals procedures were raised which should be borne in mind, especially for future systems constructed in a crisis situation. Improved communication was raised as a significant issue for public confidence that should be a legacy of the pandemic.

The findings of these discussion groups were used to design a broader survey of stakeholder views, which is reported separately in this series.

1.5 Acknowledgements

We would like to acknowledge the contribution of SQA colleagues to this report, especially Simon Allan, Ellen Macintosh, Martyn Ware and Laura Wilson. Additionally, we owe a debt of thanks to all of the stakeholders who participated in this research.

2 Perceptions of Assessment Standards in Scotland Project

2.1 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, the impetus for this research programme came from the SQA Executive, charging 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. In the Rapid Review of the National Qualification Experience conducted in 2020, attention was drawn to a perceived lack of transparency in engagement with stakeholders (Priestley et al., 2020, p4).

This report is the second in a series of four reports which look at stakeholder perceptions of standards; in this case focusing upon the findings from focus groups. The other three in this series are

1. Overview of the Perceptions of Assessment Standards in Scotland project
3. Perceptions of Assessment Standards in Scotland: Questionnaire with Stakeholders
4. Perceptions of Assessment Standards in Scotland: Employer Survey

2.2 Stakeholders' perceptions of assessment standards in Scotland

Public confidence in the qualification system is essential to maintain the currency of the certificates for the young people whose life chances depend upon them. It is vital that relevant stakeholder groups are both well-informed and consulted. 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 usefully sought. Concerns regarding assessment standards during the pandemic were an important part of the evidence-gathering. The emerging findings informed SQA's Communications Strategy and through disseminating the findings, contributed to the delivery of that strategy.

2.3 Approach to the research

Research which was at once independent, well-informed, and integrated with the operational needs of SQA was needed. Additionally, the project needed to gather information that could influence longer term assessment policy as well as the shorter-term needs for forthcoming assessment series. 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 as part of the participatory design. To ensure that the project was open to the views of participants, focus groups with 82 practitioners, lecturers, pupils, students, parents, carers, head practitioners, college principals, employers, university admissions officers and journalists were conducted first. Next, an online questionnaire survey was conducted with 918 complete responses. Finally, a telephone survey of 103 employers was conducted.

As the focus groups were conducted in advance of the 2021 summer assessment series, 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.

3 Focus Group Methodology

A participatory action research project was undertaken, involving participatory methods, dialogue across departments in SQA and rapidly-produced interim findings. The rationale for this approach, which was challenging at times, was to ensure that the research was useful to SQA and therefore benefited the public. Expertise within SQA was integrated into the design of the study and the independence of the teams in universities brought wide perspectives to the research design. Data collection techniques involved focus groups with a range of stakeholders, a survey of stakeholder views and a separate survey of employers' views. Further information about the participatory methods and the other strands of data collection can be found in the other three reports in this series.

The topic of examination standards is not new in the research literature, but there has been limited empirical research exploring stakeholder perceptions of standards and qualifications directly, particularly in Scotland. This is of note, as understandings of assessment terms are often incommensurate between assessment bodies like SQA and key stakeholders who engage with the assessment process in numerous ways.

This study focused on understanding how key stakeholders involved with the assessment and qualifications process understand the notion of standards, and how they view the fairness in a qualifications process. This research was conducted with stakeholders who engage with the qualification and assessment process at different timepoints and in different ways. Working with these stakeholders allowed us to build a holistic view of how qualifications are understood by the community.

This study used multiple methods to answer the following research questions:

- 1) What influenced the reactions of different communities to the SQA qualifications processes and results in 2020?
- 2) What factors influence different communities' perceptions of fairness for all learners in qualifications?
- 3) How do people in different communities understand the term standards in the context of qualifications? What do they believe matters?
- 4) What do people in different communities believe can be learned from experience in 2020 and 2021 in relation to:
 - a) confidence that future qualifications are fair and that standards are dependable?
 - b) actions of individuals and communities to inform public perceptions?

3.1 Data collection

The University of Glasgow team led the focus group strand of the research project; planning and organising the participation of various stakeholders. A total of 19 focus groups were conducted. Focus groups were used to ascertain participants' thoughts about and understanding of standards as well the qualifications and assessments. Focus groups were considered more useful and a more effective means of data collection than individual interviews, as the focus groups provided for the possibility of interaction and participants reacting to other perspectives in what were complex and emotive subject matters. From a pragmatic perspective, scheduling focus groups meant that we could collect data from a larger number of people and from varied groups of stakeholders in a shorter timeframe. From a participatory perspective, focus groups are one of the

most commonly used methods of data collection in participatory research. Three main strategies were employed to make the discussions as participatory as possible:

1. Sharing of broad discussion questions in advance – this allowed for participants to reflect on the questions and may increase participation from less vocal participants. It also afforded them the opportunity to offer suggestions for additional or alternative questions, in this way involving participants in the instrument design.
2. Open questions followed by probing – this allowed for a co-construction of key terms in the discussion. As a recurrent criticism of research in the assessment literature is that there is a disconnect between technical and everyday use of key terms, this ensured that participants could explicate their own definitions.
3. Iterative process of analysis and collection – the focus group discussions ended with a reflective discussion that represents the first stage of analysis as well as providing an opportunity to provide feedback that will shape future focus group discussions. In this way, participants were able to guide the direction that the research took, to at least some extent. It is key to note that the first set of focus group discussions conducted indicated that the focus group discussion approach should be continued as the participants valued the opportunity to hear from other people who were navigating the assessment and qualifications system.

3.2 Participant recruitment

Recruitment of participants represents a key moment of the research process as it is the moment where the participant can exercise their power most. The decision to participate in the study acts as a preliminary litmus test to determine whether the participant agrees that the study is worthwhile. A key challenge for the recruitment process comes about in the way in which the study is presented to potential participants. It was important that the aims and purposes as well as the methods and roles of participants were made explicitly clear from the start. This is of particular importance for a research process that claims to take a participatory approach as it promises a certain degree of decision-making power to the participant. How and to what extent participants exercise that power must be clear to them from the start. For the focus groups and surveys, participants were made aware who was conducting the research, intended goals of the project and how the findings would be used.

Where possible, those involved in preparing for Scottish Higher qualifications in 2021 were targeted for recruitment, on the grounds that they had an immediate and current interest in fairness and standards and would also have been involved in SQA qualifications in 2020, the first year of COVID restrictions. Stakeholders whose views were sought were identified in three broad categories: learners, those who deliver qualifications and users of Scottish qualifications. Recruitment of learners was therefore of secondary school S5 and S6 pupils and FE college students sitting Scottish Highers in 2021, and parents of those school pupils; those who delivered qualifications were school and FE senior managers and teaching staff organising and delivering SQA qualifications in 2021; those who use qualifications were Scottish university and FE college admissions staff; employers who used SQA qualifications in recruitment; and journalists from broadcast and print media.

To ensure that groups of participants were geographically and socio-economically inclusive, almost all participants in focus groups for learners and those who deliver qualifications were recruited through Scotland's six Regional Improvement Collaboratives (RICs) and Colleges Scotland. The RICs are structured in relation to regional needs and provide educational improvement support, fostering collaborative working across the region (Scottish Government, 2021). Acknowledging that RICs, local authorities, schools and colleges were under intense pressure at the time of recruitment, planning to deliver qualifications in the context of COVID restrictions, the research team planned to spread the load of participation as equitably as

possible. At the same time, the arrangements were intended to allow stakeholders some active involvement in determining the shape and focus of the study, consistent with the idea of participative research outlined above.

For schools, each of the six RIC Lead Officers was asked to recruit four schools from different local authorities, with a range of Scottish Index of Multiple Deprivation (SIMD) profiles, with a view to identifying 24 schools from separate local authorities and securing the agreement of their local authorities for them to take part. This agreement was important for the ethics and data protection arrangement for the project: the research team would need to know especially about any permissions or procedures in addition to those of the project team obligations; and about any issues with working online across RICs and local authorities. Once the relevant local authorities had agreed that their schools could be approached, the research team made direct contact with headteachers, who were asked to identify *either* two senior managers; *or* three Principal Teachers or Faculty Heads from three different curriculum areas or subjects; *or* two pupil and two parent participants.

Similar arrangements applied to colleges. Colleges Scotland was asked to identify and approach nine colleges from different areas of Scotland to secure their agreement to participate, after which the research team approached senior managers in the colleges directly. Three of the colleges were asked to identify two senior managers who had been integrally involved in the college's arrangements for SQA Higher qualifications in 2020 or 2021. The other six colleges were each asked to identify two lecturers or subject leaders from two subject areas who were preparing students for National Qualification Higher (SCQF level 6) qualifications in 2021. Lecturers teaching National Qualification courses that were not necessarily available in local schools were especially welcomed. These same six colleges were also asked to identify two students who were sitting Higher qualifications in 2021, preferably in the same subject areas as the lecturers and possibly from the lecturers' own classes. More mature students were particularly welcomed, to better represent the college student body.

For the users of qualifications group, colleges were also asked to identify local businesses and organisations that they worked with to provide student placements and recruitment. In the event, three of the nine colleges provided such contacts. Higher education institutions other than the larger universities that were routinely consulted by SQA were approached to nominate admissions officers to participate. Contact with media organisations was made directly by the research team.

3.3 Conduct of the focus groups

Focus groups were conducted online, using Microsoft Teams, between April and June 2021. Online focus groups were arranged, as COVID restrictions necessitated remote engagement with stakeholders. They had the distinct advantage of allowing the involvement and interactions with a wider range and number of organisations, from different parts of Scotland, than would usually be the case for in-person focus groups. We recognised, though, that the extent to which participants were able to interact effectively with each other was inevitably affected by both their confidence and experience of the online environment and the variable quality of the internet connections and equipment to which they had access. Two members of the research teams, from the Universities of Glasgow and Oxford, led each focus group, with one facilitating the discussion and the other taking notes and monitoring the use of the chat facility to ensure that comments were included in the discussion.

The learner and deliverer pairs or trios identified by schools and colleges were each invited to take part in one online one-hour focus group discussion with others in the same category from two or three different schools or colleges from different parts of Scotland, making groups of up to six participants, which was felt to be the upper limit for what would enable a reasonable exchange of different views in an online

environment (see Abrams and Gaiser, 2017). It was emphasised to RICs, Colleges Scotland and the identified schools and colleges that the project was about giving participants a voice, from their own perspectives, about fairness and standards in National Qualifications in 2020, 2021 and beyond. They were encouraged to approach participants who might especially value an opportunity to have their voice heard after the experience of 2020. It was also emphasised that the researchers were seeking to recruit participants who were not currently represented on any of SQA's groups or networks: not 'on the inside' and not regularly consulted as stakeholders by SQA.

All participants were sent in advance a Plain Language Statement describing the project and the Questions for Discussion. These were in four broad sections, relating to

- their reflections on the way qualifications results were arrived at in 2020 and public reactions to their publication;
- their perceptions of what makes for fairness in qualifications;
- their perceptions of standards in qualifications and how they are maintained over time; and
- their views on how public confidence that the 2021 results were fair and standards were being maintained could be assured.

It was intended that recruiting pairs or trios of participants from schools and colleges, rather than individuals, might support them to contribute and encourage them to discuss their views, issues and questions with colleagues and members of the school or college community in advance of the online discussion. The research team undertook to provide interim findings from the research on request to networks and learning communities as the project progressed.

3.4 Analytical strategy

Due to the participatory design, the process of analysis began during fieldwork, and continued on to an iterative process with descriptive reports forming the first stage of analysis before coding on a more analytical level. During the cycle of data collection, themes emerged that were captured in fieldnotes, which were treated as the informal analysis that takes place when in the field. This section will focus on the formal analysis that was conducted after 'leaving' the field.

Automatic transcriptions from Microsoft Teams were cleaned with reference to the recorded focus group sessions. Data was anonymised in this process, removing names of individuals and organisations. The edited transcripts were first inductively analysed, for the development of codes and for rapid reporting of draft findings for use by SQA and the project team. Summary tables were created for each focus group, showing the contribution to the discussion of individuals and the key points being made, together with researcher notes and interpretations of non-verbal communications in the recordings. The key points were also checked against field notes made by the focus group chair and facilitator and the record of the within meeting 'chat' from Teams.

At least one other researcher reviewed the resulting table for each focus group and any differences in interpretation were resolved by discussion. Once summary tables were available for all the focus group meetings for a particular stakeholder group (school pupils, college lecturers, parents, college students, school senior management team, school principals, college senior management team, employers, higher education institution admissions officers), a rapid report was produced with draft findings for that stakeholder group. The nine rapid reports were initially structured around the questions put to the focus groups but increasingly around the themes raised by participants, as the project became more mature and learned from previous rounds of reporting which themes contrasted with emerging findings from other

groups. At least two researchers reviewed the rapid draft reports against the data and any issues were resolved by discussion before the drafts were considered by the broader team or SQA staff. Feedback from team discussions informed subsequent data collection, analysis and reporting.

All focus groups were then coded manually. Words or phrases were marked as codes that could essentially be related to the research questions in a manner postulated by Elliott (2018, p.2851). These initial codes were descriptive in nature, which were then organised into more analytical themes relevant to the phase of the research. In both phases, coding was initially done by hand, and then the repeated rounds were done using CAQDAS software (NVivo 12) to have easily accessible coding frameworks. Using this software also facilitated ease of handling of the larger volume of data. NVivo 12 was largely used for the electronic management of data, however, the power of the tool to provide numerical analysis has also strengthened the internal generalisability of the data (Maxwell, 2010), when creating a more analytical code.

A mix of deductive and inductive coding was used to analyse the data with codes generated prior to data collection based on the rapid literature review that was undertaken. Coding was also adapted to the data inductively, since the principal aim of the project was to address views and concerns of stakeholders, rather than to impose preconceptions from the research literature. Coding was performed by two members of the team using the codebook. Quality checks on coding were conducted to ensure inter-rater reliability. These quality checks were conducted. The data was coded independently by both coders without conferral. This first round of independent coding was followed by a meeting where differences were discussed, the codebook was revised, and a second round of independent coding commenced (Campbell et al., 2013; Hruschka et al., 2004). The codes were compared once again, following which the analysis was distilled into the discussion that has been presented in this report.

3.5 Ethics

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.

4 Focus Group Findings

RICs, local authorities, schools and colleges were under intense pressure at the time of recruitment, planning to deliver qualifications in the context of COVID restrictions. Nonetheless, 19 of the 22 planned focus group meetings took place. Communication with one RIC was slow, resulting in the inclusion of only two schools from that area. Employers proved especially difficult to recruit, even though they had been encouraged by partner colleges to participate. It was also noticeable that there was some cross-over amongst groups. For example, several adult participants across the deliverer and user groups were also parents or family members of pupils sitting National Qualifications in 2020 or 2021 and their expressed views drew upon those experiences. Similarly, some parents were teaching in further or higher education so had views to offer based on that experience; especially about IT provision and arrangements for online learning.

4.1 Participants

Focus groups were conducted with key stakeholders involved with different parts of the assessment and qualifications process. In total, there were 19 focus groups with 82 participants, who came from 27 schools, 14 colleges, three higher education institutions and three small and medium-sized businesses and one media organisation (Table 1). Participants were from a range of local authorities across Scotland. Drop-outs in the focus groups were common due to connectivity issues, illness and other reasons. This is the reason for a single journalist being involved from the media.

Table 1 Focus group participants

Stakeholder group	Participants	Institutions	Focus groups
School senior management teams	9	6	2
School principal teachers	12	5	3
School pupils	18	9	3
College senior management teams	5	3	1
College lecturers	10	6	2
College students	8	5	2
Parents or carers	13	7	3
University admissions officers	3	3	1
Employers	3	3	1
Media (interview)	1	1	1
Total	82	48	19

As previously indicated, codes were generated both deductively (from the preceding research) and inductively (from the data). This was important to ensure that stakeholders' perceptions were contextualised, but unconstrained by preconceptions from the literature. Findings from the focus groups are presented by research question in the analysis that follows. We indicate how many focus group transcripts and excerpts across all transcripts were coded to themes, which shows the level of discussion arising in relation to each code.

4.2 What influenced the reactions of different communities to the SQA qualifications processes and results in 2020?

In this section, we examine the reaction to the qualification process in summer 2020. Codes relating to this question and the number of transcripts and excerpts coded to each are shown in Table 2. We then contrast this with the more mixed reactions to the final results and discuss some of the implications of this.

Table 2 Coding data relating to the reactions of different communities to the 2020 qualification process

Code	No. transcripts	No. excerpts
What were the different communities' reactions	11	85
Fair or unfair	2	4
Issues with clarity	7	19
Lack of confidence	4	5
Negative	10	31
Anger	4	4
Confusion	1	4
Disappointment	6	10
Loss of confidence	2	2
Unhappy	6	10
Positive	3	3
Postcode lottery	8	13
Reaction to the media	6	8
Reflection of students' effort	2	2

4.2.1 How and why did communities react to the qualifications process

Most participants considered that the shift to teacher-inferred grades in summer 2020 probably represented the best solution for a difficult situation. Nonetheless, there were extremely negative reactions to the way in which the process had been handled. Participants reported anger, frustration, disappointment with a process that had felt, to them, characterised by consistent uncertainty, change and poor communication. This uncertainty and confused communication led to a loss in confidence in the SQA as an institution:

I think if you chop and change things based on elements of pressure or whatever else, it will always raise eyebrows and lead to more questions being asked ...

(College Lecturer)

This sentiment was primarily directed at SQA, with other organisations, such as Education Scotland, noticeably absent from discussions. This was the case even when issues fell outside of the remit of SQA. Most participants reflected that the shift to teacher-inferred grades probably represented the best solution for a difficult situation.

An overwhelming feeling across all participant groups was one of poor communication and lack of clarity. Participants reported losing confidence in SQA due to the untimely nature of communication and in some instances an overload of information, making it hard to distil the important parts of a communication:

The SQA routinely put out messages at 7.30 at night - Friday evenings have been a particular favourite - that happened quite frequently when they were moving the goal posts earlier on in the year. And now we're getting bombarded with almost daily messages from the SQA. Obviously before Easter the message was that this was the window where the assessment should be happening. Now they've realized, oh, that's going to mean pupils are facing an awful lot of assessment and there's a backlash against that. And they seem to be backpedalling from that. And again, the statements that we get, day after day, I seem to be reading them, and I think there's a lack of clarity around what we are now expected to do.

(School Teacher)

Participants reported that difficulties in reacting to last-minute and poorly structured communication were exacerbated by the fact that the public often received information at the same time, giving them no time to plan their responses when parents and pupils asked for clarifications. Alongside this, participants felt that the communication from SQA did not always align with messages in the media and from politicians. In all, these factors combined to undermine the effectiveness and credibility of SQA communications, leading in many instances to frustration, disappointment and anger.

4.2.2 *The role of the media*

Integral to this breakdown in confidence was the role of the media, who were perceived as contributing to the negativity surrounding the process. Social media, as well as mainstream media, was seen to have played a role.

I think the comments that were in the media, in my opinion, were quite flippant, because they had no idea of the effort and the time and the expertise that went behind that.

(College Lecturer)

This 'postcode lottery' thing down-grading everybody. I think that was really wrong. I've got a news article loaded up right now: a lassie got five As in her prelims. And then after [moderation] she got two As and three Bs. So, at that point the SQA hasn't listened to the teacher or lecturer at all.

(College Student)

Reference to the 'postcode lottery' highlighted both the negative reaction to the process and the role that the media had in fuelling it. Many participants used this phrase, which was a repeated refrain in both the media and in political rhetoric, relating to the use of prior attainment data in the statistical moderation process that was initially used to produce results in 2020. Importantly, the reaction to this was driven to a large degree by the perception that the moderation process was inherently *unfair*. Many participants felt that the moderation that SQA had undertaken had disproportionately schools and pupils in disadvantaged areas:

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

(School Parent)

Our interview with a journalist indicated that the narrative was led by the young people themselves. Personal stories were of interest to the readership of the publication and the fact that young people had been

protesting publicly when they had been subjected to prolonged period of lockdowns and when there was still concern about spreading of the virus was newsworthy. Grade inflation, which would be of interest in a normal year, was seen as a subsidiary narrative compared with changing the teacher grades and of course the subsequent policy reversal.

I think that for us, I mean because of our audience, for us the fluctuation in the pass rate, obviously that was a really big story, but I do still think, you know, that the proportion of teachers' grades that were changed, that's what we thought would be of most interest to our readers.

(Journalist)

Importantly, some of the school and college SMT (senior management team), lectures and teachers pointed out that the system was not necessarily creating new inequalities, rather that there was greater awareness of them this year than others. For example, the use of achievement distributions as part of the moderation process happens every year, but only this year was it flagged as *unfair*. This mirrored a recognition from some participants that the public 'uproar' was not necessarily a rational response to the process. These participants reflected that had the process started with teacher judgements then probably that would still have resulted in uproar causing the opposite reversal:

I think that what might have happened if it was going down that route [teacher-inferred grades from the start] is that the government may have come in for a lot of criticism when the results came out, there would have been a lot of talk about grade inflation, which inevitably would be there. And I think there have been questions asked about it, so it was probably almost in a no-win position.

(School SMT)

I agree that moderation needed to take place. I think that had that the algorithm not been applied, you know, the media would have found cases where somebody had graded way higher than everybody else and made a big whirlwind out of that and we would have all been just as upset had it not applied as had it applied.

(Employer)

Participants then recognised that the negativity surrounding the process, while directed predominantly at SQA, was not entirely their fault. In particular, participants felt that better communication, both with schools and colleges as well as with the media more broadly may have mitigated much of the negative fallout from the process.

4.2.3 Reactions to the final results

In contrast to the *process*, participants' perceptions of the final results were more balanced. Many participants reflected that these probably represented the best solution to an incredibly challenging situation. Generally, participants were happy with the final set of results – although this raised the question of how sustainable a system can be with this level of grade inflation.

I initially felt quite disappointed when the results first came out and seeing that a lot of people were possibly downgraded due to their postcode which kind of ruins the whole standardised thing of doing assessments, but was delighted to see that the SQA went back and changed what they did and rethought the results they gave out to people.

(School pupil)

Parents and school-based participants felt happy with the final results, often seeing it as the fairest outcome to a challenging year:

Obviously, I was pleased with it

(School parent)

... going from heads down to really quite delighted at the end of the day

(School SMT)

In part this was because the grades achieved were significantly higher than in previous years at a national level. College lecturers and SMTs, on the other hand, often seemed quite angry:

... I was absolutely furious with what happened to the grades because we have like an astronomical amount of quality assurance in place at the college to make sure that there was no overestimation.

(College Lecturer)

I was actually more upset when the Scottish Government got involved and then they went all back to teacher results because I think the results that we finished up with in that year were unrealistic and as a receiving body, I know I have students in [Higher National] classes who have got a B in Higher English who cannot write, who have teachers, you know, who were put under pressure by parents in a way that the college sector is not. And I think schools exaggerated results and that's why SQA made the changes that they made.

(College Lecturer)

This tended to stem from a sense that the final results predominantly favoured schools and was reflective of a lingering disillusionment with a system which focusses more heavily on schools over colleges. College students and college pupils were mixed in their reactions. Some reported a sense of disappointment and failure when results were first released, followed by relief when they were adjusted back to the teacher-inferred grades. Others worried that the grades awarded were perhaps not a fair reflection of the cohort and some pupils may have over-achieved:

Well yes and no. I guess to be fair, my grades last year were great and I was very happy with my grades, but I thought that it was very lenient and I guess you know everyone sort of knew what grade boundary they were sitting at. And some people who were maybe six to seven, you know, sort of sitting on the passing boundary, they did pass and might not have actually passed on the day.

(School Pupil)

Some pupils spoke about peers or who had made a big jump from prelims to final results. They suggested this was unfair and evidence of favouritism:

We had different view. We thought that it was fair, but also unfair. For example, like we feel like in our school, quite a lot of people were upgraded in subjects and places that they maybe shouldn't have been as much as they were. So, for example, some people got a C in the prelim, but then ended up with an A in the exam, and we also feel like that's kind of a bit of favouritism,

(School Pupil)

Finally, employers were ambivalent overall. They did, however, recognise that the increase in the number of A grades being awarded was a red flag. Employers mentioned that they had placed less emphasis on qualifications even prior to COVID and saw other things as being important for their recruitment process, such as teamwork, soft skills, IT and so on.

4.2.4 What does transparency mean in this context?

A repeated theme was the desire for ‘transparency’ as well as ‘certainty’. This represents a challenging tension and contradiction as clearly the SQA was faced by rapidly changing societal circumstances, making ‘certainty’ unobtainable. Much of the dialogue around the pandemic has pointed out that the pandemic has highlighted or illustrated existing systemic flaws and inequalities, rather than creating new ones. Similarly, the pandemic increased the public’s awareness and focus on the qualification system. The inequalities that caused outrage over summer 2020 were pre-existent, but suddenly they were exposed to greater public awareness.

When the examination process was relatively obscure, people tended to have high levels of trust in the judgements made. However, the process of making assessment less obscure meant that people started to see some of the complexities of the process and lose trust in it. The relationship between transparency and trust is therefore complex.

4.3 What factors influence different communities’ perceptions of fairness for all learners in qualifications?

It was important to first understand how people in different communities conceptualised fairness in relation to its conceptualisation in the literature on qualification on standards. Table 3 gives the definitions of fairness that were used as deductive codes for this part of the analysis (drawn from Nisbet and Shaw, 2019; 2020). The conceptual mapping of *formal, implied contractual, relational, and retributive* has been categorised further into technical and ethical fairness. Ideas of technical fairness pertain to whether the rules were followed (formal), including contractual expectations, even if the rules or contracts were implied or inferred (implied contractual). Ethical fairness rests on more social notions, rather than conforming with process or regulation. In ethical fairness, ideas of social justice, such as treating alike cases similarly (relational) or even taking into account what has gone before the outcome was produced (retributive).

Table 3 Definitions of fairness coded, from Nisbet and Shaw 2019, 2020

	Conceptualisation of fairness	Definition
Technical Fairness	Formal Fairness	Denoting accuracy or appropriate application of a rule or design.
	Implied contractual	Something is fair if it meets the legitimate expectations of those affected.
Ethical Fairness	Retributive fairness	In which outcome is regarded as fair if it is an appropriate reward (or penalty) for what has gone before.
	Relational fairness	Treating (relevantly) like cases alike.

Tension between notions of ethical fairness and technical fairness were palpable in the data. It also became clear that fairness is conceptualised differently by the different communities and that a different level of importance and emphasis is attributed to the concept by the different communities. As Table 4 shows, Table 4 shows, there was an extremely high emphasis on ideas associated with the notions of **ethical fairness** in the focus group transcripts, emphasised most by teachers and lecturers. It is interesting to note that students' definitions of fairness resonated more with the ideas of **technical fairness**; The analysis highlighted that for lecturers, senior leaders, and teachers who were involved in the day-to-day dealings with qualifications and assessments, higher emphases were placed on ideas that were more closely related to the definitions of ethical fairness, where students, higher education institution admissions officers, and employers' ideas were most closely linked to notions of technical fairness.

Table 4 Coding data relating to the factors influencing perceptions of fairness

	No. Transcripts	No. Excerpts
Technical Fairness	8	26
Ethical Fairness	13	111
Implied Contractual Fairness	2	7

Table 5 shows some of the factors highlighted across all the communities as to what needs to be true to consider qualifications and assessment as fair.

Table 5 Coding data on factors which make an assessment fair

	No. Transcripts	No. Excerpts
Type of assessment	9	27
Clarity	6	14
Expertise	5	12
Consistency at a systems level	5	12
Assessment outcome	5	7

Two key factors that were highlighted across most communities as important in building the feeling of 'fairness' were knowledge and transparency. This was less obvious for employers who emphasised more strongly that fairness in qualifications depended on consistency. For school SMT members, most indicated that having knowledge of how the process of assessments were graded, or even how qualifications were structured were helpful in making qualifications and their assessment 'feel' fair. The role of 'felt' fairness or 'perceived' fairness in qualifications is not uncommon nowadays, particularly in the context of COVID-19.

I think the key for me is - what makes the results feel fair and I think knowledge is the thing that allows it to feel fair.

If you know what is happening to your piece of work as a learner and that - you know the process and the rigour, then you feel that you're being treated fairly, and I think the end that your result is a fair representation of the work that you've done.

(College SMT)

A particularly key driver that was discussed in detail across all the communities was assessment format. This was felt particularly strongly amongst students who felt that the high stakes examination approach was unfair to different learners and their needs, and the anxiety caused by the nature of examinations was detrimental to a true understanding of a learner's understanding.

I think, well, I know certainly because I'm dyslexic. I know people, other people who are dyslexic. I don't think making them sit in an exam room is very fair. 'Cause yes, you get extra time and things like that. But that's just making you sit, and stare at a piece of paper for longer and.... That - for me, certainly, that was hopeless. 'Cause I'm one of those people who are hands on and very practical. But exams are very academic. They're not made for practically minded people, and I don't think that it was right - when they give you extra time, it's to help you. But actually, it can make the whole situation a lot worse, a lot more stressful.

(Student)

Higher education admissions officers indicated that one of the ways that this might be addressed was through standardised assessments throughout the year, an idea that was echoed by employers and parents, who felt that measures such as the teacher grading could be partial, leading to feelings of unfairness. Further, college lecturers noted that the evidential basis for grading was important and college SMT discussions also noted the importance of external verification of judgements for standardisation. The role of local authorities in helping to standardise teacher assessments across schools was raised by members of school SMTs. Principal teachers mentioned that training for teachers in marking and provision of grade boundaries for use with the question papers would have improved fairness.

Standardisation in the form of the same examination for everyone across the country being taken at the same time and under the same conditions was raised as a basis for fairness by college students, though the counter-arguments were also identified. Having anxiety, a bad day or for those students for whom examinations were not suitable would not necessarily experience the standardisation that examinations produce as fair. Application of a national standard was seen to influence perceptions of fairness by college SMT members. School teachers and college lecturers indicated that

[SQA] could not 'devise a system that is fair for all students'

(College lecturer)

Even with the case of examinations, participants pointed out that there were some students who would just 'nail an exam' without putting in much work throughout the year.

The conditions during the pandemic, in which schools and colleges had not all been able to cover the entire curriculum and with variable degrees of lost learning amongst pupils and students, meant that standardisation was not seen as feasible by parents. Schools and colleges had adopted different approaches to collecting assessment data, as befitted their circumstances, in the academic year 2020-2021. Parents reported that young people were not always given adequate warning that assessments, including examinations, would be made, which meant that they could not prepare at times.

It is also interesting to note that for some of the participants of the focus groups, particularly members of college SMTs and school Principal Teachers, it was difficult to decouple the notions of standards and fairness, highlighting how intrinsically the ideas are linked to one another.

I think we're going to come to the standards later on. You were saying that the comparison of this year and last year, it's incomparable. You've got to hold your hands up and say this is not comparable. There's absolutely no comparison to my son's experience of Highers compared to my

daughter's four years ago experience of Highers. It's totally different, it could not be any more different, unpredictably different, exceptional. ... it's so exceptional that, you know, we kid ourselves on that it's the same standard. I mean, you're in university, I'm in university too, I teach in university. There's absolutely no way that the degree that I taught this year was the same as the degree that I taught two years ago. None at all. Absolutely none at all. And I think it's been heroic on their students' part to hang on in there and get through it. It really has been.

(Parent)

Across communities, there was an acknowledgement that societal structures in general create a level of unfairness for learners and the qualifications process itself cannot counter inequalities that exist in terms of access to resources and materials that would give some learners an advantage over others. That said there was a call across some of the respondents, particularly parents and employers, that there needs to be mechanisms put in place so that the system can create a degree of moderation to address some of the inequities that may play out in results.

We live in an extremely unfair society and many people are in positions of privilege and many people aren't. And I mean I think that often these results are, these inequalities are reflected in results, opportunities for young people after their qualified. So, I think you know there's a wider debate around fairness in our society. I think in terms of the exams, um, these concerns are apparent - these inequalities or these unfairnesses are apparent, to me, from things that happened previous to last year and it's it is not really just teachers who are responsible for addressing these inequalities, but there must be some kind of system where there are, where there can be a moderation that does not penalize people because of their social and economic circumstances or their ethnicity

(Employer)

Different stakeholder groups viewed fairness primarily from either an ethical perspective or a technical perspective based on where they sat within and interacted with the qualifications system. Parents and learners (who are the most directly impacted by the awarding of a certain grade), for example raised matters that were more closely attributed to ideas around ethical fairness, where practitioners and to an extent employers were more concerned with ensuring that technical fairness is maintained. This tension is one that highlights that there is a constant need for communication across the different stakeholders in the system to understand where each's understanding may be stemming from.

4.4 How do people in different communities understand the term standards in the context of qualifications?

Table 6 lists the definitions used in coding the focus group data, following the classification of these in Baird et al. (2018). The literature on the meaning of standards in national qualifications has proposed a number of ways of thinking over the years. Baird (2018) organised the definitions in an ecological model, according to the level of explanation that each offered. Examinee-level definitions indicate that the object of interest is something related to the examinee. The only definition in this set of data which is at examinee level is the **criterion-referenced** definition (Popham and Husek, 1969); this is where students' outcomes are determined by reference to pre-specified performance criteria, which are written and normally shared in advance of the assessment. These criteria would outline what students must know and be able to do, to be awarded the grade. Thus, the criterion-referenced definition refers to examinee performance. Baird et al. (2018) outlined three different assessment paradigms, with distinctive ways of thinking about assessment broadly and standard setting in particular. Outcomes-based assessments are associated with vocational and higher education; very often these have criterion-referenced standard setting systems. A **cohort-referenced** definition (William, 1996) refers to a system in which a specified proportion of the candidates are awarded

each grade, which is typically the same from year to year. For example, that 10% of candidates should be awarded a grade A. This definition concerns the management of examination systems rather than being related to factors associated with the examinee. Curriculum-based, high stakes national examination systems sometimes use this approach to standard setting. Georgia, South Korea, South Africa and the state of Victoria in Australia have had exam systems using cohort referencing (Baird et al, 2018, p302, Table 14.6). A **due process** definition (Cizek, 1993) is again a systems level definition and relates to the notion that there should be a transparent adherence to the rules and procedures regarding grading. A further level of definitions is the systemic level (Baird et al., 2018), in which exam boards use an umbrella definition that refers to multiple levels. **Attainment referencing** (Newton, 2011; following Baird, Cresswell and Newton, 2001) is a definition in which both examinee performances and statistics relating to those performances are taken into account when setting standards. This is the official approach to standard setting for Higher examinations in Scotland, as well as for A-levels in England and the Leaving Certificate examinations in Ireland. This approach allows for changes in difficulty of the examination to be adjusted for in the standard setting, given that it will have affected the apparent quality of students' performances. Therefore, a hard exam set one year would not result in lower grades due to the fault of the exam; and vice versa. Unlike the criterion-referenced approach, this definition is concerned with students' underlying attainment rather than their performances alone. A final approach to defining standards emerging from the focus group data in this study relates to the **currency** of the qualifications. This is the value of the grades, in terms of access to education and employment. Currency of qualifications is a systemic level definition which goes beyond the exam board, since it encompasses articulation with the wider education system and with the labour market. This approach does not speak directly to examinee characteristics or to statistics, but the currency of qualifications could still depend upon these factors.

Table 6 Definitions of standards coded

Level	Definition	Paradigm	Informs us regarding
Examinee	Criterion-referenced	Outcomes-based	Outcomes related to pre-determined performance criteria
System	Cohort-referenced	Curriculum-based	Standing within the population taking the assessment, since the proportions being awarded the grades remains the same from year to year
System	Due process	All	Grading following the application of pre-specified rules and procedures
Systemic	Attainment referencing	Curriculum-based	What students know and can do, taking into account changes in difficulty of the assessment
Systemic	Currency	All	Value of a qualification grade

Most of the data relating to the understanding of the term standards related to definitions of standards (Table 7). Aspects of assessment which fitted with notions of different assessment paradigms and timing of evidence collection also arose, but were more minor issues in the data and therefore are not discussed below. As Table 7 shows, most of the discussion about what standards meant used a criterion-referenced approach, followed by a cohort-referenced definition. SQA's official approach to setting standards, attainment referencing, was not often used when participants discussed standards. Of course, participants did not always use the terms in Table 6 directly, but the way they discussed standards was consistent with these definitions and coded accordingly.

Table 7 Coding data relating to understanding of standards

Code	Transcripts	Excerpts
Definition of standards	14	65
Criterion-referenced	25	44
Cohort-referenced	9	16
Due process	4	6
Currency	4	4
Attainment-referenced	2	3
Other definition	8	22

The first code in Table 7 contains data on the general discussion of standards. Much of the text related to features of how the standards were set in 2020, including the use of moderation (which came to be known as the algorithm), the timing of data collection, aspects of the data used to set standards in schools, relationships with SQA and so on. Alternative systems were sometimes proposed, such as,

I think they should do like an average of the student's work on like assessments that you've done in exam conditions, because you might have an off-day in one of them and get like a worse mark than you usually would. But then sometimes you just do really well, almost unexpectedly. That'd be like really great if that happened over like the exams the normal years. But since it could be if you only took that, it could be seen as like less of a representation of you. So, if they considered more of an average of all that you've done in throughout the year it would be a better representation of the student as a whole.

(Pupil)

Participants in the focus groups were asked what they understood by the term 'standards', as related to qualifications. Data related to this part of the focus group discussion, or elsewhere in the discussion were coded with regard to how people discussed the meaning of standards. Use of terms did not always mean that participants supported that way of thinking, only that they discussed standards in this way.

A criterion-referenced approach to standards was most frequently mentioned; the criteria required for grades were raised, such as in the following quotation,

The standards are obviously the guidelines that they're 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)

Participants spoke of guidelines, levels of knowledge and understanding required, the Scottish Qualifications Credit Framework (SQCF) and were positive about criterion referencing, together with moderation or external verification. A need to keep these standards relevant to the needs of business was mentioned by college lecturers and college students discussed the importance of standards for educational progression. Often participants spoke about the minimum standards required to gain a grade and recognised that, within a grade, pupils might have different performances but still be worthy of the grade overall, as the following quotation shows,

I know that in my subject, for instance, I have seen many native speakers achieve an A quite reasonably and without too much of an effort and another candidate ... getting the same grade is not the same achievement. ... So, I do accept that they will have different skills, different abilities within the same grade, but there is something that the standard here is there to show a minimum.

(School teacher)

Many participants understood that cohort referencing was a system-level definition, due to small samples in schools and colleges making this not a viable proposition at that level. For example, one participant said,

... if you've got a large cohort, you would expect again the law of averages should kick in that you'd have is consistent sort of bell curve type set of results, but that isn't going to be the case where you've got small cohorts of children. They might all add up nationally to the same number of kids year in year out. So, one of our subjects has run about 15,000 pupils approximately each year who sit the subject. But I know some schools where they might, they might be putting in 90 pupils pretty much a year for that subject and others where it's 7... You know if you got a small cohort, their results can be amazing one year - 100% pass rate and another year you can have, you know, as we've experienced ourselves as low as 12%.

(School teacher)

Although a cohort-referencing definition was less frequently used as a basis for the discussion of standards, those who did use it were positive about it and recognised the need for some system-level checks upon the distribution of outcomes. Higher education officers in particular saw the need for expected patterns of achievement to be reached, year on year and saw this as the responsibility of SQA. College lecturers also indicated that radical changes in outcomes at a national level in any one year would be surprising. Grade inflation was mentioned by a number of participants. In cases where criterion referencing was discussed, participants often considered that the statistics should not affect whether pupils were awarded the grades. Many of the discussions recognised the difficulties inherent in the task of setting standards and some related the use of cohort referencing to the use of the algorithm.

Due process definitions of standards were sometimes alluded to. For example, a member of a school SMT said,

They're about setting, I suppose, partially setting the rules, ... on how we conduct assessments, and then about the grade related criteria. But I think it's a combination of all of that, so not just agreed criteria for the actual assessments, but for the conduct of them how we apply them.

(School leader)

Participants also discussed the way in which the algorithm had been applied and how data had been collected to evidence the school grading in 2020, but we did not code that text as related to due process because it was specific to the year in hand, rather than a discussion about what standards meant in general.

In the three instances using an attainment-referenced definition, participants stated that both criteria and statistics were used in standard setting to overcome variability in the difficulty of the assessments, but the text was unclear in each case, reflecting the complexity of this issue. For example, one participant said

... if there was a particularly nasty paper, they'd lower the percentage thing. I think the point being made it's on your group as a whole's intelligence is ... but at the end of the day I think it's just them trying to make papers that are somewhat equal in difficulty ...

(School leader)

The definition of standards as currency was not often alluded to, but one participant used the term directly, as follows,

I think it's a bit like a currency, basically. The currency has to maintain its value over time. So, if someone's got a Higher English this year, you kind of hope that essentially, it's very much like the Higher English you got last year, ... the year before.

(School leader)

4.5 What do people in different communities believe can be learned from experiences in 2020 and 2021 in relation to fairness, standards and public perceptions?

Overall, the main message from participants across the focus group discussions was that the disruption due to the COVID-19 pandemic both represented an opportunity and highlighted the need to move away from an over-reliance on high stakes standardised examinations at the end of a sequence of learning. While the qualifications processes of 2020 and 2021 were far from ideal, they represented emergency responses, rather than intentional and planned changes. Participants welcomed the opportunity to re-examine the ways in which qualifications are awarded if this was supported by consultation with key stakeholders in a thorough and un-rushed fashion. This section is divided into two halves. The first deals with participants' perceptions of the future of fairness and standards and the second with the various roles and actions that participants highlighted as important to a fairer qualification system in the future. Each of these issues were well-represented in the data (Table 8).

We'd also probably like to take away some of that flexibility in assessment design, and really begin to really explore that a little bit further what makes a good assessment. I think, COVID to an extent exposed inherent unfairness in all assessment systems, and we can only do things that can mitigate the risk to individual students, but absolutely taking forward some of that flexibility [...] We've learned a lot and we can take that forward to make a much better and fairer assessment system for national qualifications.

(College SMT)

Table 8 Coding data relating to what can be learned from experiences in 2020 and 2021

Name	No. Transcripts	No. Excerpts
Learning for the future	18	158
Future qualifications are fair and standards are dependable	18	76
Informing public perceptions in the future	14	59
Other recommendations for the future	8	23

4.5.1 What can be learned to ensure future qualifications are fair and that standards are dependable?

4.5.1.1 Role of the final exam

If we didn't do a single exam at the end of the year as before but SQA regulated, say, three throughout the year, you'd get a real picture of the student. And I think teacher input is a good thing because it makes the person seem like more of a person instead of just a percentage on a piece of paper

(School Pupil)

Participants felt that currently the qualification system was over-reliant on final examinations. This over-reliance resulted in the system's inability to cope with shocks, as demonstrated by the 2020 qualification process. Alongside this, participants felt that examinations measured only a limited set of skills, and didn't always reflect the full range of a learner's capabilities. Finally, while examinations were considered 'fair' in that every pupil sat the same paper under the same conditions, many participants felt they were inequitable, in that they did not reflect the different circumstances of learners they were seen as susceptible to learners having a good or bad day. Overall, participants felt that a broader and more flexible process, possibly regulated and moderated at the level of the RIC or the local authority, would represent a fairer and more dependable process for qualifications.

One of the lessons that I think we should learn is that exams and tests aren't the be all and end all. [...] My point is that a good qualification system should have a range, like a healthy diet. And I think that when we put our eggs all in too many narrow baskets, that's when we can get unstuck.
(School SMT)

While participants recognised the limitations of relying too heavily on a final examination, there was little desire to get rid of exams altogether. Most participants suggested it should serve as a piece of additional evidence, amongst a range of others. Some participants were in favour of a return to modular examinations, whereas others felt that course and teacher assessment could be important components of a broader process. One participant in the employers group suggested that a final examination could serve as a diagnostic tool for teacher-assigned grades.

Only one participant, the journalist, reflected upon the reasons for more emphasis on final examinations rather than continuous assessment, despite Curriculum for Excellence.

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

4.5.1.2 *The appeals process*

There was a sense of disappointment and anger from some participants regarding the appeals process in 2020, that stemmed from a sense of powerlessness due to the perceived degradation of the process. Participants felt that the appeals process should be school-led and evidence-based, with schools making appeals and having to provide evidence to the SQA to support their case. Some school SMT recognised the potential for parents to cause problems if they felt their child had been graded unfairly and so wanted the assurance of support from RICs or the SQA in the event of a dispute.

I think there has to be a way for a child to have a bad day and still get the grade that they deserve. If they've worked really, really hard all year, telling them that they don't deserve an A or B or even a C or a D in some instances is not particularly fair
(School teacher)

4.5.2 *What actions are required to inform public perceptions of future qualifications*

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

Participants felt that the public had lost a lot of confidence in the process of qualifications and worried that this would mean that they would lose confidence in the qualifications themselves. This would result grades that were awarded in 2020 or 2021 being perceived as less valuable. Participants were unanimous in their opposition to this and felt that, in the short-term everyone involved in the process, and in particular employers and universities, needed to publicly validate grades:

Schools themselves, educational authorities, employers, universities, they've got to speak up and say, this was a very serious event for the whole of society. For schools, for colleges, universities, businesses, people, individuals, all ages, and it had an epic impact and school pupils were no different from that. I think schools and education authorities have got to speak up, against anybody that might say, 'Oh no, there's no way that was as good as when I was a pupil'. There's no way you went through anywhere near what these kids are going through. So, the schools and education authorities have got to let people know that you're doing the best you can to assess the quality of the pupils that are coming out of the system and going into the colleges and work and all the rest of it. It was exceptional times and the schools should be telling people that, and education authorities should be telling people that, so there's not any kind of scepticism surrounding any of the COVID generation of pupils that are going to be coming out in the last two or three years, and, well, the next six to seven years actually, because we're talking about the kids there who are with them right now. So, it's behoven on their education authorities and their own schools to really get the message out there, you know, these pupils did a sterling job, you know, under exceptional circumstances. We are proud of them.

(School parent)

In the longer term, participants felt that improved communication and consultation processes were needed, both to develop fairer assessment mechanisms and to communicate these to schools and the wider public. Participants pointed to the perceived failure of SQA to engage effectively with the public and stressed the importance of learning from this to have better dissemination in the future. Across the focus group discussions is the repeated tension between need to act fast and desire to slow things down so that everyone can be informed, and so that decisions can be made in a considered and transparent fashion. Participants felt that, to develop and sustain a broader and better qualifications process, in depth consultation with stakeholders would be necessary:

This is an hour out of our day and it's so valuable. I know what 6 to 7 schools are doing. It took one hour. I mean that that would have been a whole RIC. That's all the SQA had to do, have meetings like this with head teachers over a week and by the end of the week they would have had a national picture and we would have felt like we've had our say.

(School SMT)

4.6 Conclusions

Qualitative data findings suggested that participants considered a wide variety of issues related to assessment standards but the most prevalent approach to defining standards was criterion referencing. With

SQA's long history in vocational education through its predecessor organisation, SCOTVEC, it is perhaps not surprising that this definition is most widely used and understood. However, criterion referencing was introduced to Scottish qualifications generally in the 1980s and 1990s (Devine, 1988; Brown, 1991), just as it was introduced to academic qualifications globally to elucidate the qualitative meaning of standards (e.g., see Sizmur and Sainsbury, 1997). Reliance on the expertise of the examiners and, indeed trust in the teaching profession more generally to set standards is a cultural aspect of education in Scotland. Nonetheless, this placed the use of a statistical moderation system, which became known as the algorithm, in opposition to the prevailing understanding of assessment standards amongst stakeholders.

Overall, participant suggested a more flexible and decentralised approach to the qualifications process would be fairest and most dependable. However, there remained many tensions within these discussions, which we will briefly mention here. These include, teacher bias, teacher workload and capacity and the dominance of 'assessment'.

Participants seemed to suggest that teachers should lead on, or at least influence, the awarding of qualifications, because teachers know the learners best and are therefore in the best position to award the fairest grades. While participants acknowledged the possibility of teacher bias influencing this process, the majority felt this could be overcome with rigorous moderation and sufficient evidencing. However, the issue of teacher bias is one that will require careful consideration should the qualifications process move towards including an element of teacher assessment.

Carrying out continuous assessment, alongside the moderation required for a system which relies more on teacher assessments, represents a significant increase in teacher workload. Over the last two years, teachers felt that their workload had increased significantly due to the increased burden or their role in the qualification process, with one teacher commenting about the 2021 process, as follows,

I would say from our perspective, in music and art, we have worked tirelessly. I couldn't give you a number of hours that we've put in extra to not only gather evidence, maintain the standard that's set that you expect of assessment and then cross-mark everything.

(School teacher)

This increase in workload was coupled with a recognition of the needed for training. One of the consequences of the pandemic on education has been the renewed focus and interest in assessment. Teachers have both improved their assessment literacy and recognised the limitations in their understanding. This makes for a particularly fertile professional development environment. If the system is to move towards one which includes teacher assessment as part of the qualifications process, it will be important to both ensure there is the necessary time for teachers to carry out this role effectively, as well as providing sufficient access to training opportunities for teachers to improve their assessment literacy and understanding of standards.

Finally, there is an interesting and important contradiction between the idea that people are doing too many assessments and that each assessment carries too much weight. Participants felt that learners should do fewer assessments, because assessment takes time, is stressful, and can distract from learning. However, they also suggested that the one-shot exam carries too much weight and moving to 'chunked' (modular) exam systems would reduce the stake of each individual assessment. Navigating this contradiction, between reducing the stakes of each individual assessment without increasing the number of assessments such that they dominate even more than currently, will be an essential element of a successful qualifications process in the future.

While a sample of 82 stakeholders represents a large sample of qualitative data, it is of course still a small sample of stakeholders' views in Scotland. Thus, the findings in this report elucidate stakeholders' perceptions but our conclusions regarding their prevalence are limited by the nature of the research. As such, these findings were used as the basis of the design of a questionnaire study, which followed the focus group research.

5 References

- Abrams, K.M. and Gaiser, T.J. (2017) Online Focus Groups. In *The SAGE Handbook of Online Research Methods*. 435 – 449. DOI: <http://dx.doi.org/10.4135/9781473957992.n25>
- Baird, J., Isaacs, T., Opposs, D. and Gray, L. (2018). *Examination standards: How measures and meanings differ around the world*. UCL IOE Press.
- Baird, J. (2018). Chapter 14 - The meaning of national examination standards. In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (Editors) *Examination standards: How measures and meanings differ around the world*. UCL IOE Press. 2 – 25.
- 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.
- Brown, S. (1991) Assessment within the Scottish education system. *Journal of Curriculum Studies*, 23:6, 539-544, DOI: [10.1080/0022027910230604](https://doi.org/10.1080/0022027910230604). <https://doi.org/10.1080/0022027910230604>
- Campbell, J.L., Quincy, C., Osserman, J., Pedersen, O.K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42, 294–320. <https://doi.org/10.1177/0049124113500475>
- 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](https://doi.org/10.1080/02671522.2012.754231)
- Cizek, G.J. (1993) ‘Reconsidering standards and criteria’. *Journal of Educational Measurement*, 30 (2), 93–106.
- Devine, M. (1988) Grade-related criteria: the Scottish experience. In H. Black and W. Dockrell (Eds.) *New Developments in Educational Assessment*. Edinburgh, Scottish Academic Press.
- Elliott, V. (2018). Thinking about the Coding Process in Qualitative Data Analysis. *The Qualitative Report*, 23(11), 2850-2861. <https://doi.org/10.46743/2160-3715/2018.3560>
- He, Q. Boyle, A. and Opposs, D. (2011) Public perceptions of reliability in examination results in England. *Evaluation & Research in Education*, 24:4, 255-283, DOI: [10.1080/09500790.2011.629722](https://doi.org/10.1080/09500790.2011.629722)
- Hruschka, D.J., Schwartz, D., St John, D.C., Picone-Decaro, E., Jenkins, R. A., Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, 16, 307–331. <https://doi.org/10.1177/1525822X04266540>
- Maxwell, J.A. (2010). Using Numbers in Qualitative Research. *Qualitative Inquiry*, 16(6), 475–482. <https://doi.org/10.1177/1077800410364740>
- Newton, P.N. (2011) A level pass rates and the enduring myth of norm referencing. *Research Matters*, Special Issue 2, 20 –26.

- 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](https://doi.org/10.1080/0969594X.2019.1586643)
- Nisbet, I. and Shaw, S.D. (2020) *Is Assessment Fair?* Sage Publications Ltd.
- Popham, W.J., & 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.
- Scottish Government (2021) Review of the Regional Improvement Collaboratives. 16 December. <https://www.gov.scot/publications/review-regional-improvement-collaboratives/>
- 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](https://doi.org/10.1080/03054985.2012.760264)
- Sizmur, S. and Sainsbury, M. (1997) Criterion Referencing and the Meaning of National Curriculum Assessment. *British Journal of Educational Studies*, 45:2, 123-140, <https://doi.org/10.1111/1467-8527.00041>
- William, D. (1996) 'Standards in examinations: A matter of trust?'. *Curriculum Journal*, 7 (3), 293–306.