College of Policing
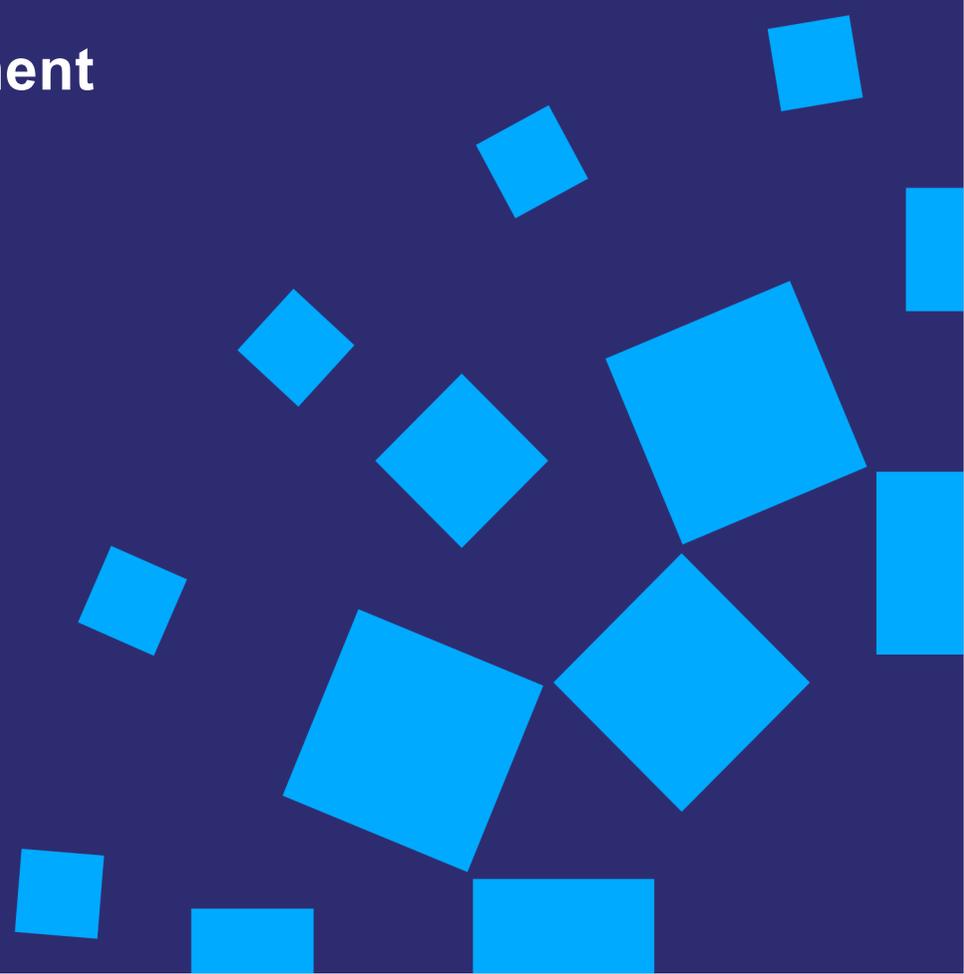
Working together
to share knowledge

Evaluation of the online assessment process for police recruitment

Final report

April 2021

college.police.uk

Tribute to a dear friend and colleague

In publishing this report we pay tribute to our dear friend, colleague and lead researcher for this evaluation Jenny Kodz. Jenny sadly passed away on 26 January 2021. Our thoughts and deepest sympathy are with Jenny's husband, Barry and daughter, Pei Gui.

Jenny joined the Civil Service in 2002 and was with the College in its predecessor form as the NPIA. Jenny was an incredibly diligent and well-respected researcher, totally dedicated to her field. During her career, Jenny contributed to many areas of research, particularly improving diversity in policing, which Jenny cared passionately about.

Everyone who knew Jenny will remember her kindness, grace, warmth and that she always had time for others, to support, to talk, to laugh. Jenny shared her wisdom gladly, ever keen to mentor those embarking on their research careers. We will miss her greatly and each have our fond memories.

Research too provides us with a collective memory, about what happened and why, considerations to inform our decision making and most importantly valuable learning we can use now, and in the future, to continually enhance how we do things – creating a lasting impact. The development of the online assessment process and Jenny's evaluation of it will definitely do that, feeding into, and contributing to, the major programme of change taking place through the Police Uplift Programme and its legacy.

Contents

Summary

Background

The online assessment process for new recruits was designed and developed as an emergency interim measure to support policing to continue to recruit during the COVID-19 pandemic. The requirement for this method remains and the College will continue to deliver online assessments, subject to further enhancements, to sustain recruitment until the end of March 2022. Alongside this, the College is working with the Police Uplift Programme and forces to ensure that other, locally managed stages of the recruitment process are fair, effective and support diverse recruitment.

Assessment results

- Overall, 72.4% of candidates were successful compared to 71.3% for SEARCH® (the previous assessment process).

- The percentage of White, female and male candidates who were successful at the online assessment process was very similar to the success rates of these groups previously at SEARCH®.

- More ethnic minority candidates (excluding White minorities) were successful with the online assessment process (62.6%) compared to SEARCH® (55.8%), which has improved the adverse impact ratio (85.1%).

- However, there was still a gap between the success rate of White (73.6%) and ethnic minority (excluding White minorities) (62.6%) and there was still adverse impact among some specific ethnic groups. Further work is being undertaken to better understand this disparity.

Assessment process

- The inter-rater reliability study showed that the consistency in scoring was fair-moderate, with scoring for the written exercise more consistent than the competency-based interview and briefing.

- Assessors appeared to be less confident about their assessments for the competency-based interview and briefing exercise and overall wanted more training and support.

- The majority of candidates (84%) were satisfied or very satisfied with the delivery of the online assessment, a

similar proportion to SEARCH® and Day One. Feedback from candidates suggested that the exercises may not have given them an opportunity to demonstrate their full range of skills.

Feedback from stakeholders

- Overall, stakeholders recognised that the College had designed and delivered an online assessment process that was essentially fit for purpose.

- Delivered at pace, the online assessments have enabled forces to continue recruitment, ensuring that a critical need for the service has been met.

- The pandemic was felt to have been a catalyst to modernise the selection and assessment process, removing blockages to change. There was no strong desire among stakeholders to return to SEARCH® or Day One but there was support for including more interactive elements online.

- Some initial issues with delivery had caused problems in forces but guidance and processes had been updated by the College in response to feedback.

- Forces used to managing their own assessment centres felt that moving to a force-led model would provide increased control and flexibility. Some stakeholders suggested, at the time of the interview, that prior to any commitment on the future model, further information from the College was required on costs, delivery and understanding adverse impact.

1.   Introduction

1.1. Online assessment process for police recruitment

As soon as COVID-19 restrictions were implemented nationally and face-to-face assessment centres had to be paused, the College of Policing worked closely with the Police Uplift Programme (PUP) to develop an online police recruit assessment process at pace. This interim model enabled forces to continue recruiting police constables during the first phase of the COVID-19 pandemic and continues to do so.

These purely online assessments were introduced as a provisional solution during the pandemic in the place of the face-to-face SEARCH® assessment centre and the rollout of Day One. Day One is the face-to-face assessment for police constables, which had been piloted as the replacement for SEARCH®. However, the planned further rollout was paused due to the COVID-19 pandemic.

The online police recruit assessment process consists of four exercises: a situational judgement test (SJT), competency-based interview, briefing exercise and written exercise. Each of the exercises is designed to make assessments against competencies and values set out in the College of Policing Competency and Values Framework (CVF), as follows:

Competencies assessed	Exercise		
	Competency-based interview	Written exercise	Briefing exercise
Integrity	✓		
Public service	✓		✓
Transparency	✓		
Impartiality		✓	
We are innovative and open-minded	✓		✓
We are collaborative		✓	
We take ownership	✓		✓
We analyse critically		✓	
We deliver, support and inspire		✓	
We are emotionally aware			✓

The SJT is electronically marked automatically and all the other exercises are marked on a five-point scale by assessors, who were contracted by the College during the initial phase of implementation when the evaluation took place. The process is supplemented by a bespoke modular programme of mandatory online training for assessors, which

aims to provide assessors with the skills and knowledge required to ensure all candidates are being assessed fairly and effectively. The training covers the process, the roles and responsibilities of the assessor, assessor best practice, biases and barriers to effective assessment, assessment principles and the assessment criteria. Assessors also undertake specific exercise training, which covers assessment methodology, using the online system and the exercise content. At the end of the training the assessor is then asked to complete a benchmark exercise, providing them with practice of using the online system along with marking against the assessment criteria.

Implementation of the online assessments

PUP worked closely with forces to establish precise recruitment needs and early adopter forces were selected, who then worked with the College and PUP to test and benchmark the online exercises. A detailed wider implementation plan was developed using prioritisation data from PUP and the College provided support to forces during the implementation process, which covered all 43 Home Office forces by November 2020.

The online assessment process was hosted on a third-party platform, LaunchPad. LaunchPad worked with existing technology to automate the recruitment process, using video interviewing technology to capture candidates' responses. The exercises were entirely asynchronous, pre-recorded without any live interaction between candidate and assessor. Alternative video platforms that could provide face-to-face assessment were considered but did not meet security requirements.

When the online assessment process was first introduced, the link between LaunchPad and a force's applicant tracking system (ATS) was made through a manual process. The administration of candidate applications has now been partly automated by integrating the College's Assessment Information Management System (AIMS) with LaunchPad to allow for a simple transition and tracking of applications.

1.2. Evaluation

1.2.1. Aims

The evaluation was designed and conducted by College researchers and the College Selection and Assessment team.

The evaluation aimed to:

- inform decisions about the online process

- identify learning for incorporating technology into the recruit assessment process in the longer term

Questions the evaluation sought to explore were:

- Do assessment outcomes show any differences between minority groups and the majority group that are attributable to the online process?

- How satisfied are candidates and assessors with the online assessment process?

- Are assessors able to assess candidate performance consistently and accurately?

- What are senior stakeholders', regional and force leads' views on the implementation of the online assessments?

- What are the perceived advantages/disadvantages of central and local force administration of the process?

- What learning can be drawn from the introduction of online assessments in forces?

1.2.2. Methods

To address these questions, the following mixed-methods approach was adopted.

- Data analysis of assessment outcomes, including diversity outcomes

- Candidate and assessor feedback surveys

- Inter-rater reliability study

- Interviews with regional leads, force leads and senior stakeholders

Alongside this evaluation, the College also commissioned occupational psychologists at the Fairway Group to undertake an independent peer review of the online assessment process, with a specific focus on fair selection and inclusion. The review concluded that the selection exercises and materials were valid and objective, incorporating many

aspects of best practice, and which provide clear guidance for candidates, assessors and quality assurers. The review is published separately on the **College website**.

Data analysis of assessment outcomes

The College undertakes ongoing analysis of the online assessment process outcomes. This report presents outcome data from 17,206 candidates who participated in the online assessment process from 4 May 2020 to 14 January 2021.

Success rates have been analysed by protected characteristics[1] with an adverse impact calculation conducted for each of the following: age, disability, ethnicity, religion and sex. A selection process is said to have adverse impact against minority or underrepresented groups when the pass rate for that group is significantly below that of the majority group. The adverse impact ratio (AIR) is calculated by dividing the minority group percentage pass rate by the majority group pass rate. A commonly used convention is the four-fifths or

80% rule, which suggests that a measure may have adverse impact if the AIR is lower than 0.8 or 80%, although the ambition is for the online assessment process to deliver no difference in success rates. Throughout the report, the term adverse impact is used when the AIR is below 80%. Conversely, if the AIR is above 80% no adverse impact is reported.

Candidate assessor feedback surveys

A candidate feedback survey was designed to explore experiences and perceptions of the online assessment process. The survey was open for candidates to complete between 16 September 2020 and 2 October 2020. The College distributed the survey to candidates via the recruitment leads of 17 forces. These were forces that had gone live with an online assessment process between 10 August 2020 and 16 September 2020. Forces were asked to send the survey to a sample of approximately 5,390 candidates who had completed or attempted the process

---

[1] The **Equality Act 2010 (legislation.gov.uk)** protects against discrimination based on nine characteristics.

within the previous month. This figure was based on the number of candidates who were expected to go through the process for their representative force during this period.

Descriptive statistics and comparisons are typically reported for the quantitative responses. Some statistical significance testing of the responses by ethnicity has been undertaken and is stated where applicable. Free text responses were thematically analysed and coded by two researchers.

A survey was also undertaken to explore assessors' experience of the online assessment process and training. The survey was sent to all 607 assessors and responses were received between 16 September 2020 and 2 October 2020. Data was analysed and descriptive statistics reported.

Inter-rater reliability study

An inter-rater reliability (IRR) study was completed in October 2020 to estimate the extent to which online assessors were able to assess the performance of the same set of candidates consistently and accurately.

The study looked at consistency of assessment scores when different assessors assessed the same candidates. The study focused on three exercises: the written exercise, briefing exercise and competency-based interview. The SJT was not included because this was automatically marked. The following process was undertaken.

- For each of the three exercises, 90 to 100 randomly selected candidates (cases) were assessed three times by three different assessors and the three assessment scores were compared for consistency.

- The three assessments for each case were the live (original assessment) and two re-marks conducted for the purposes of the study.

- The re-marking was conducted by 10 randomly selected assessors (30 in total) for each of the three exercises.

- The IRR was assessed using a one-way inter-class correlation (ICC). The results were measured on a scale from 0 (no agreement) to 1 (perfect agreement). The interpretation of the IRR values is taken from the Landis and Koch scale, shown below.

ICC statistic	Interpretation
<0.00	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

All interviews were audio recorded, transcribed and coded into themes. The aim was to draw out common views, where perceptions of different individuals were consistent with each other, as well as insightful exceptions.

Stakeholder interviews

To understand how the online assessment process has been implemented and received by forces, as well as what lessons have been learned, 23 in-depth stakeholder interviews were undertaken between August and October 2020.

These included interviews with:

- 11 force recruitment leads
- Seven regional force leads for PUP
- Five senior stakeholders

2.   Online assessment process success rates

Between 4 May 2020 and 14 January 2021, 72.4% of 17,206 candidates who participated in the online assessment process were successful. The pass rate for SEARCH® was 71.3% between 31 August 2015 and 29 February 2020.

The following charts and tables show success rates by ethnicity, sex, age, disability and religion. Full data tables that present success rates and adverse impact for each characteristic, including specific ethnic group categories, are included in Annex B.

Ethnicity

More ethnic minority[2] (excluding White minorities) candidates have been successful with the online assessment process compared to SEARCH®, whereas the success rate for White candidates has remained the same (see chart below). The

AIR for the online assessment process for ethnic minority candidates was 85.1%, which is an improvement compared to 76.0% for SEARCH® and above the 80% threshold at which adverse impact is suggested to have occurred.

Percentage of ethnic minority and White candidates successful with the online assessment process and SEARCH®
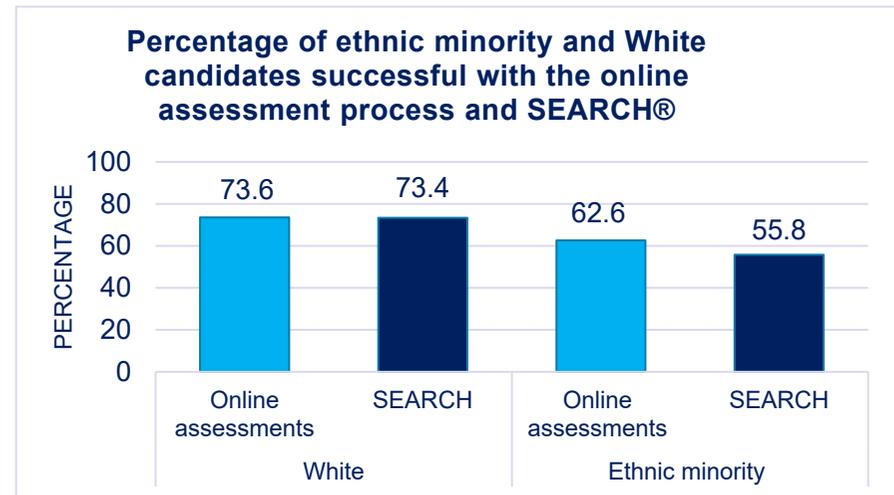
Improved success rates were evident in each ethnic minority group (see chart below). However, despite notable increases in successful candidates among Asian and Black ethnic groups as a whole, the AIRs for these ethnic groups was

[2] The comparison made is the White ethnic group against all other ethnic groups combined. The White group contains White minorities, such as Gypsy, Roma and Irish Traveller groups.

slightly under the 80% threshold (78.7% and 79.0%). The remaining ethnic groups were all above 80% (see table opposite).

Percentage of successful candidates with the online assessment process and SEARCH® in each ethnic group
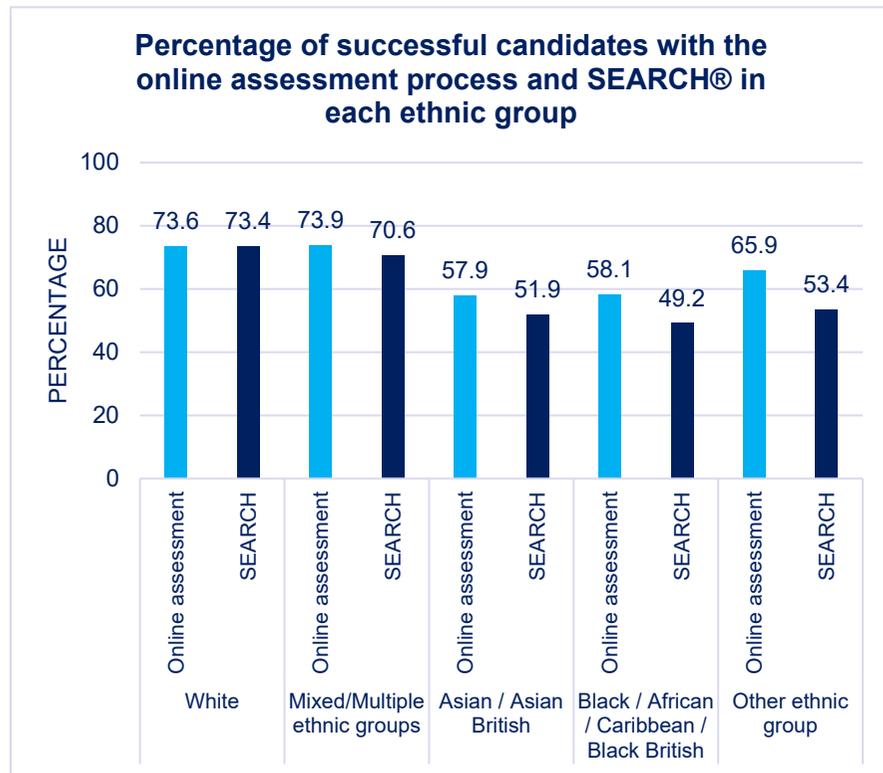
Ethnicity	Mixed/ multiple ethnic groups	Asian/Asian British	Black/African/ Caribbean/ Black British	Other ethnic group
AIR (%)	100.5	78.7	79.0	89.5

However, within the Asian and Black categories as a whole, there were different outcomes among candidates from specific ethnic groups. Indian (AIR = 87.7%) and Black Caribbean (AIR = 90.3%) candidates were above the 80% adverse impact threshold, whereas Bangladeshi (AIR = 66.2%) and Black African (AIR = 64.4%) were below (see Annex B).

While the introduction of the online assessment process has led to an improvement in the success of ethnic minority candidates compared to SEARCH®, a gap between the success rates of White (73.6%) and ethnic minority candidates (62.6%) remains. This was particularly evident for Asian (57.9%) and Black (58.1%) ethnic groups overall.

This gap was also demonstrated in the comparison of ethnic minority and White outcomes for the individual exercises of the online assessment process (see chart below). Each of these differences in White and ethnic minority exercise scores is statistically significant, although the effect size, which is an indication of the size of the relationship, is small/small-medium and lower than that at SEARCH® (see Annex B).

Online assessment exercise performance for ethnic minority and White candidates

Exercise	Group	Mean percentage point score
Written	Ethnic minority	55.8
Written	White	60.1
Briefing	Ethnic minority	57.8
Briefing	White	62.2
CBI	Ethnic minority	61.3
CBI	White	63.4
SJT	Ethnic minority	86.2
SJT	White	87.3
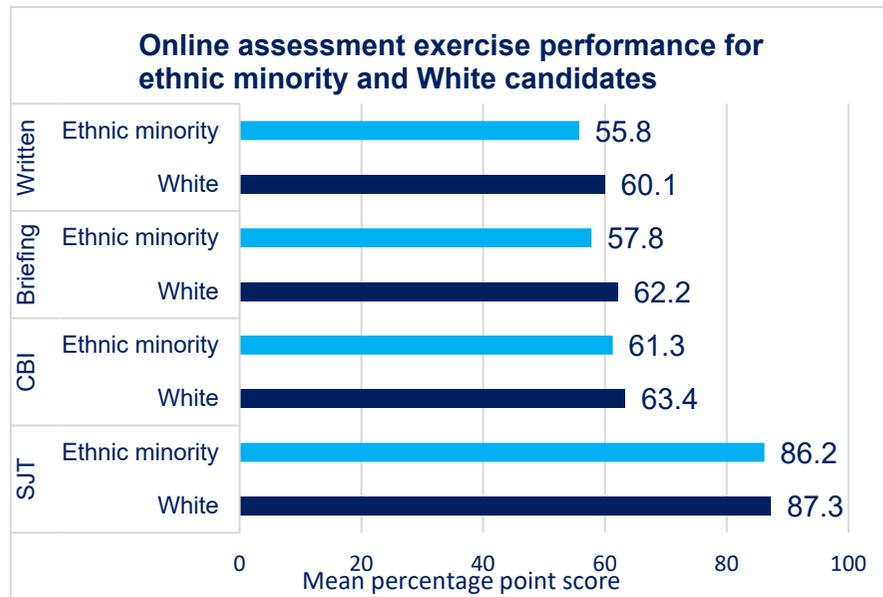
Sex

There were no changes in the success rates of female or male candidates from SEARCH® to the online assessment process. Female candidates (75.3%) continued to outperform male candidates (70.0%) (see chart below). As males form the majority group, there was no adverse impact (AIR = 107.6%).

Percentage of male and female candidates successful with the online assessment process and SEARCH®

	Female		Male	
	Online assessments	SEARCH	Online assessments	SEARCH
PERCENTAGE	75.3	75.7	70	69.2
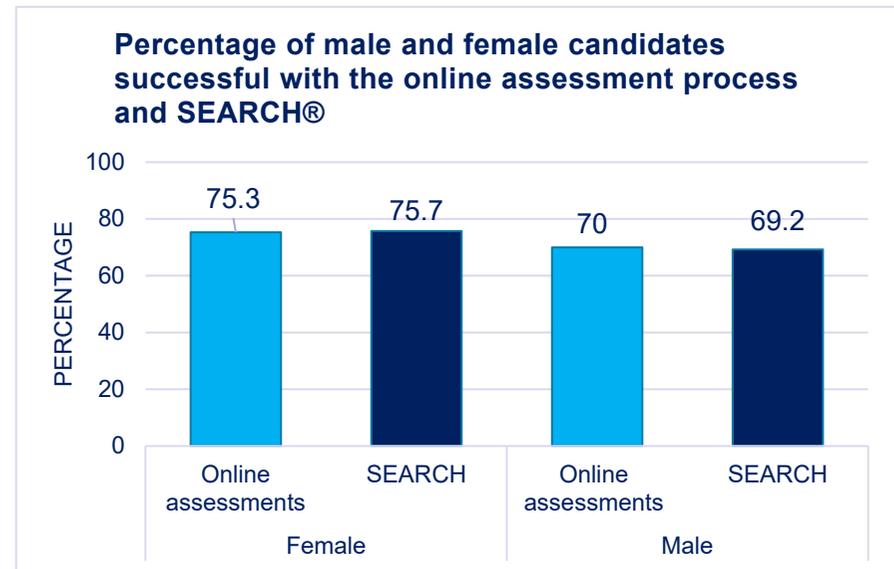
Age

Candidates aged under 26 had the highest success rate for the online assessment process (71.6%), an increase in performance from SEARCH® (60.9%) (see chart below). The success rate of candidates aged between 41 and 55 decreased at the online assessment process (69.4%) compared to SEARCH® (75.8%) but there was no adverse impact (AIR = 93.7%).

Percentage of candidates by age successful with the online assessment process and SEARCH®

71.6   60.9   74   72.6   69.4   75.8

Online assessments   SEARCH   Online assessments   SEARCH   Online assessments   SEARCH

Under 26   26-40   41-55
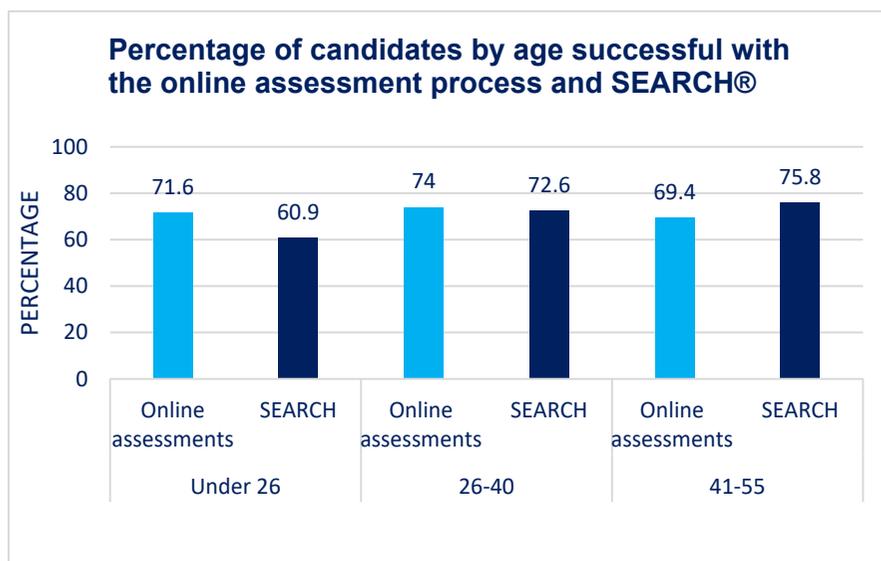
PERCENTAGE

Disability

The percentage of successful candidates at the online assessment process was slightly higher for those without a disability (72.0% compared to 70.2%) and there was little change from SEARCH® (see chart below). There was no adverse impact for disability (AIR = 97.5%).

Percentage of candidates with and without a disability successful with the online assessment process and SEARCH®

70.2   69.1   72   71.3

Online assessments   SEARCH   Online assessments   SEARCH

Disability   No disability

PERCENTAGE
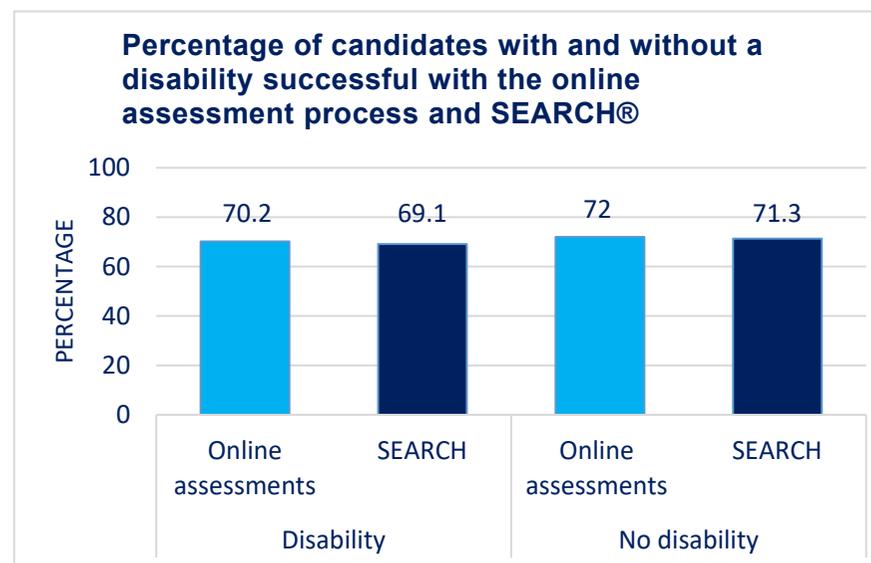
Religion

For all but one religious group, success rates were higher with the online assessment process compared to SEARCH® (see chart below). Buddhist candidates had a slightly higher success rate for SEARCH® (69.7%) than online assessments (67.3%). Adverse impact was found for Muslim (AIR = 77.1%) and Hindu (AIR = 77.8%) religious groups but this had improved from SEARCH® (69.7% and 75.8%).

Percentage of candidates by religion successful with the online assessment process and SEARCH®
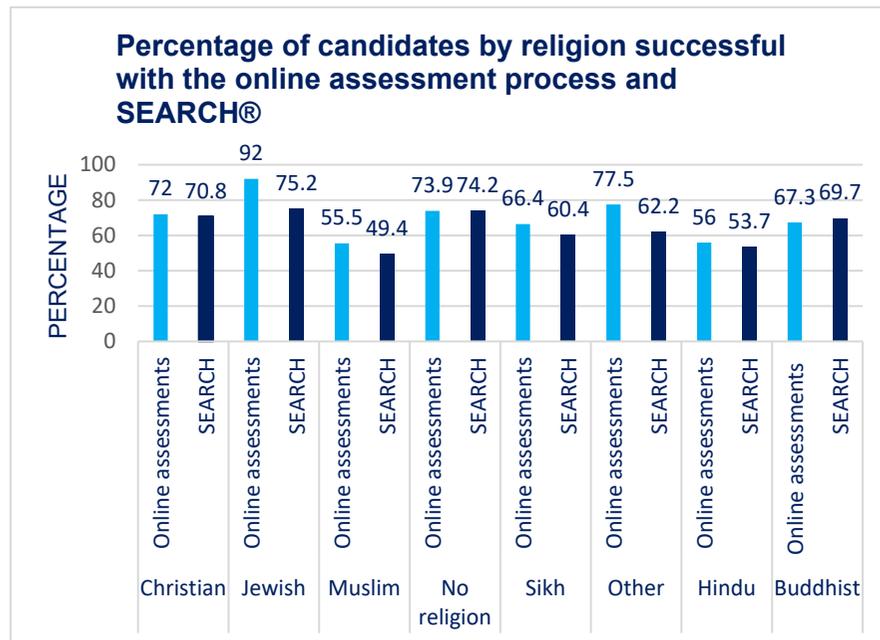
3.   Candidate and assessor perceptions

3.1. The candidate survey

A survey seeking feedback on experiences and perceptions of the online assessment process was responded to by 931 candidates. This response represented approximately 10% of all candidates who had participated in the online assessment process at the time of the survey (September 2020).

3.2. Candidates' demographics and assessment journey

Demographics

Of the 931 respondents[3]:

- 92% stated that they were White and 7% from an ethnic minority[4] (excluding White minorities)

- 41% of respondents were female, 59% male

- 57% were aged under 26, 38% aged between 26 and 40 and 6% over 41

- 7% stated that they had a disability

- 29% had at least one parent/guardian with a degree-level qualification

See Annex C for tables presenting full respondent demographics.

Assessment completion and withdrawals

Most (92%) candidates stated that they had started the assessment process between 1 June 2020 and 30 September 2020. For some this may have been the date of their application to a force, not the start of the online process. Survey respondents applied to 16 different forces overall.

Of the candidates that responded to the survey, 99% said they had completed the assessment process. Of those, 67%

---

[3] Due to rounding, totals may not equal 100.

[4] The comparison made is the White ethnic group against all other ethnic groups combined. The White group contains White minorities, such as Gypsy, Roma and Irish Traveller groups.

had received their assessment result, with 75% successful. Only 13 candidates had not completed the assessment process. This may suggest that candidates who withdraw from the process may be less likely to respond to a survey asking for feedback. Reasons the 13 candidates gave for withdrawing from the process included technical issues, a change in work or personal circumstances and disliking the online process due to it feeling impersonal.

3.3. Candidates' perceptions of the online process

Overall satisfaction

The survey asked a series of questions relating to overall satisfaction with the online assessment process. Some of the questions in the survey were also included in previous candidate surveys of the SEARCH® and Day One assessment centres[5]. Where possible, comparisons with the earlier findings are included here, as well as breakdowns of key questions by ethnicity, gender and disability. A summary of free text comments on the process are also presented to provide further insight in the sections to which they are relevant.

- 84% of respondents were satisfied or very satisfied with the overall delivery of the online process. This compares with 88% of Day One and 86% of SEARCH® candidate survey respondents.

- Breakdowns of the online assessment candidate survey responses showed that 79% of ethnic minority candidates (n=67) were satisfied or very satisfied compared to 84% of White candidates. Further analysis using a Mann-Whitney U test has shown that this difference was not statistically significant ($U = 27610$, $p = .570$) (see Annex C). This suggests ethnic minority candidates were not more likely to be dissatisfied than White candidates.

---

[5] Day One and SEARCH® survey comparisons reported in College of Policing (2019) Day One Pilot Assessment Centre interim evaluation report with recommendations version 1.2 [Available on request].

- Female candidates were very slightly more satisfied than male (85% compared to 83%).

- A lower proportion (70%) of candidates with a disability (61 candidates in total) were satisfied or very satisfied with the overall delivery of the process than candidates with no disability (85%).

- 82% of respondents said they would be likely to recommend the online assessment process to a friend or colleague. This compares with 79% of SEARCH® and 87% of Day One survey respondents.

More candidates noted positive than negative views in their free text comments about the overall process. Respondents who found the overall experience positive reported that the assessment process was enjoyable, user-friendly, well organised and in a professional format. The few negative comments included perceptions that the process was unfair, impersonal or poorly organised.

Comparisons candidates made with face-to-face assessments

338 survey respondents had previous experience of a face-to-face assessment centre (in any sector). Of these candidates, 46% agreed that undertaking the assessment exercises online enhanced their experience compared to face to face, 27% disagreed and 27% were neutral.

Reflections on the differences between face-to-face and online assessments were the most frequently mentioned theme in responses to open-ended questions in the survey. Some candidates noted the benefits of the online assessments in terms of enabling recruitment to continue during the pandemic. Candidates who reported that they favoured the online process considered it more relaxed yet still professional. Those who raised concerns about the online assessments talked about the impersonal nature of the process and the lack of interaction with an assessor preventing any clarifying questions or development of rapport between the candidate and the assessor. Some felt that they could not express themselves fully online or that the assessor

could not gain a full understanding of them and see their potential for the role.

Provision of information prior to assessment

Most (82%) online assessment candidate survey respondents agreed that the candidate guide provided enough information to support preparation for the online assessment process. Only 9% of respondents disagreed with this statement.

Some survey respondents reported that the information sent to candidates prior to undertaking the online assessment was comprehensive and helpful in the free text responses to the survey. Others felt there was a lack of clarity around timescales for the whole process to be completed and did not feel sufficiently informed of what to expect. These respondents provided feedback on communication issues they experienced early on in delivery, such as a requirement for more specific information for different exercises or highlighting faulty links in emails. Some candidates also reported having little time to prepare and too much information to digest.

Suggestions for improvement included the provision of mock or sample questions, the provision of example screenshots of the layout of different screens and exercises prior to the assessment and the provision of a forum or 'live chat' facility.

Feedback on the exercises

Over 80% of candidates perceived each of the exercises in the assessment process to be relevant to the role of a police constable. The following proportions of candidate survey respondents felt the exercises were relevant or very relevant:

- 94% written exercise
- 92% briefing exercise
- 89% SJT
- 85% competency-based interview

From the free text survey comments, it is evident that the briefing exercise was the element that caused most concern, with more observations about this exercise than the other three. Some candidates felt that the questions in the briefing exercise were confusing, that they needed more time and that prior policing knowledge was required. Some felt their responses did not reflect what they would have actually done

in a real-life situation. An opportunity to interact with the assessor or a role-play element was a suggestion for improvement. Also for the interview, some candidates commented that the opportunity to interact via a platform such as Zoom or face to face would have been preferable.

Views on aspects of the assessment process

Candidates were asked a series of statements about the assessment process (see table opposite). Just over half (55%) of candidates agreed or strongly agreed that the online assessment process allowed them to demonstrate their full range of skills and abilities (21% disagreed). The percentage who agreed or strongly agreed was lower than among Day One candidates (76%) and slightly lower than SEARCH® (59%). A greater percentage (63%) of ethnic minority candidates agreed that they could demonstrate their full abilities compared to White candidates (54%), although this difference was not statistically significant (U = 25719.500, p = .139) (see Annex C).

92% of candidate survey respondents agreed or strongly agreed that the LaunchPad technology was easy to use. Some technical issues were mentioned in free text comments

including information not displaying correctly on the screen, particularly on tablets and mobile devices, and issues with connectivity or the system crashing.

Statement	% of candidates strongly agree/agree
The LaunchPad technology for the online assessments was easy to use	92%
The information provided during each exercise was clear	90%
I had enough time to complete the exercise	85%
I received thorough communication and support throughout the online assessment process	77%
The online assessment process allowed me to demonstrate my full range of skills and abilities	55%

There were positive comments about the technical support provided throughout the process. But for a few candidates, technical difficulties were a distraction that they felt prevented them performing to their full potential.

In relation to information provided during the exercises, 90% of respondents felt the information they received was clear. A

lower proportion of respondents (77%) agreed or strongly agreed that they received thorough communication and support throughout the online assessment process.

There were comments from candidates about receiving too much and also about too little information. Some candidates commented on the early communication issues and fed back that the volume of emails and links had been confusing. However, at the end of the process, some candidates were concerned that they did not receive a confirmation email that their application had been received.

Most survey respondents agreed or strongly agreed that they had sufficient time to complete the exercises (85%). Timing issues reported in free text responses were mostly based around wanting more preparation time before demonstrating competencies. More time to record responses would have also been beneficial to some respondents.

Of those that reported having a disability, 59% were satisfied with adjustments to meet needs, 9% very unsatisfied and 19%

neutral. Some with learning disabilities expressed through the free text responses that they had trouble reading and hearing at the same time.

3.4. The assessor survey

253 assessors responded to a survey to provide feedback on their experiences of the online assessment process and the assessor training (a response rate of 42%).

Assessors' demographics and assessment experience

Of the assessors who responded to the survey[6]:

- 88% stated that they were White, 10% from an ethnic minority and 2% preferred not to say

- 50% of respondents were female, 49% male

- 55% of assessors were over 55 years old, 38% were 41-55 and 5% were under 40.

---

[6] Due to rounding, totals may not equal 100.

At the time of the survey, 59% had spent 11 or more days assessing in this process and 38% less than 10 days. 58% were new assessors and 42% had previous College assessor training experience.

Assessors' perceptions of the process

86% of assessor respondents agreed or strongly agreed that the exercises enabled candidates to demonstrate evidence relevant to the competencies.

When thinking about the whole experience of assessing candidates online, 76% of assessor respondents said they would be likely or very likely to recommend the online assessment process to a colleague. This percentage was higher than was found in response to a similar question in a Day One assessor survey (55%) but lower than among SEARCH® assessors (91%). However, the number of survey respondents to the SEARCH® (151 respondents) and Day One (38 respondents) surveys was lower.

In free text comments at the end of the survey, similar suggestions were made to those from candidate survey respondents about opportunities for interaction between the candidate and the assessor. Another recommendation was the inclusion of scenarios or role plays. Some concerns were mentioned about the security of the assessments.

Support for assessors

Assessors were asked how confident they felt in assessing candidates correctly after completing the training for each of the exercises. Most assessors felt confident about assessing each exercise, but were most confident with the written exercise (92% confident) compared to the briefing exercise (84% confident) and the interview (81% confident).

As shown in the table below, online assessors were less satisfied with the support available to them as compared with assessors surveyed for SEARCH® and Day One. However, most online assessment assessors (87%) found the LaunchPad technology easy to use.

Free text responses at the end of the survey included a suggestion that assessors would have found some feedback on their scoring beneficial, especially when new to the role. Some were concerned that assessors may not be marking consistently.

Statement	% of assessors strongly agree/agree		
	Online recruit n=253	SEARCH® n=151	Day One n=38
I had enough support to successfully perform in my role as an assessor	59%	94%	68%
The technology at the assessment centre was easy to use	87%	43%	76%

Other suggestions included more timely email responses to queries and a dedicated phone helpline for assessors. Assessor comments on the assessor training included requests for more personal guidance, more practice examples and a reduced time lag between assessor training and live assessments. Other comments made by assessors included difficulties accessing the technology to revisit the training or not knowing that it was possible to do so.

4.   Inter-rater reliability (IRR) study

Overall results

The results of the IRR suggest that the consistency in marking the three exercises was 'fair' for both the competency-based interview and briefing exercise and 'moderate' for the written exercise. Overall, the findings indicate a higher level of consistency and IRR for the written exercise.

Exercise	IRR	95% confidence interval	n	Interpretation
Competency-based interview	0.29	(0.16, 0.42)	93	Fair
Written exercise	0.54	(0.43, 0.64)	100	Moderate
Briefing exercise	0.34	(0.21, 0.47)	92	Fair

As an illustration of the consistency of marking, the proportion of cases where the three markers agreed on a score, and the proportion of times they differed (on the assessment scale of 1 to 5) is shown below.

Exercise	Perfect agreement (%)	Differ by one grade (%)	Differ by more than one grade (%)
Competency-based interview	37%	49%	14%
Written exercise	51%	44%	5%
Briefing exercise	44%	46%	10%

These findings reflect survey responses which suggested that assessors were more confident in the written exercise compared to the CBI and briefing. The absolute agreement for the written exercise (51%) is similar to that of A-Level marking for essay based exams, such as English or History, as reported by Ofqual in 2018 (Marking consistency metrics, 18/6449/2).

More detailed analysis of each exercise can be found in Annex D.

Alternative scoring methods

An alternative IRR method was tested, using the average marks for the competencies that the assessors look for within the exercises (eg, integrity or public service in the interview) rather than the single overall mark for the exercise itself (such as the briefing, and presented in the analysis above). This method produced IRR scores that were the same or marginally above the standard marking scheme, indicating that the two sets of marking were relatively well aligned with each other. It should be borne in mind that some of the increase in IRR will be down to the increased granularity of using average scores (ie, a continuous scale compared to a ranking from 1 to 5).

The College will explore alternative scoring methods further and undertake work to understand the effect on pass mark boundaries and adverse impact for minority groups.

Further discussion on the alternative IRR methodology can also be found in Annex D.

5. Interviews with stakeholders

5.1. Overall views

Overall, stakeholders were positive about the delivery of the online assessment process, highlighting the speed and efforts of College staff. Through the interviews, the evaluation also sought to identify lessons to improve future ways of working. These focused on resourcing, initial delivery issues in forces, project management, costs and the role of the College in future delivery.

Force stakeholders were consistent in recognising the role of the College in implementing the online assessment process. They felt that the College had delivered, at pace, a system that enabled forces to progress police officer recruitment, meet PUP targets and maintain a stable workforce. Furthermore, some suggested that without the online assessment process, a prolonged interruption to recruitment would have begun to impact on the capability of forces to provide essential services. It would also have left a huge recruitment backlog that would be disruptive for higher

education institutions planning for admissions on to the Police Constable Degree Apprenticeship.

> "The College delivered a system that enabled forces to continue recruiting all the way through COVID. Not just for the 20,000 but to maintain a stable workforce – we lose people every day. They would have been severely understaffed if they couldn't continue recruiting."
>
> **[Stakeholder interview]**

Benefits and risks of the online assessment process

Stakeholders highlighted a number of common benefits of the online assessment process, aside from meeting the urgent need to respond to the COVID-19 pandemic. They felt that for candidates an online assessment was a flexible, convenient and modern process more equivalent to other professions. Forces would benefit from a capacity to process higher volumes of applicants and respond dynamically to demand, while freeing up force estates.

Stakeholders also identified associated risks they felt required monitoring or managing, but did not view these as serious enough to withdraw the online assessment process. They suggested that the integrity of the online process could be open to potential abuse as the process was unsupervised and the identity of those completing the assessments could not be validated with total certainty. This raised concern that exercises might be completed by an individual who was not the named candidate. However, stakeholders were satisfied that the risk could be managed and candidate identity verified.

Force stakeholders welcomed early and promising data around adverse impact but felt ongoing analysis was required to monitor and understand how minority groups performed. There was also concern that some specific reasonable adjustments were not available in the early days of the online assessment process.

The use of a third-party platform had created a database of future police officers and it was felt this created a potential security risk – but mitigations were again felt to be in place.

5.2. Attraction and support

National focus through Uplift

The drive to recruit 20,000 additional officers through PUP has introduced renewed focus on workforce diversity at the national and local level. A national recruitment strategy has been developed, new governance structures established and a marketing campaign, managed through the programme, launched in September 2019. Senior national stakeholders commented that the national campaign, the first in several years, had been successful in reaching a high volume of potential candidates through national and social media, but more progress was needed to improve workforce diversity.

The increased recruitment through Uplift has also presented forces with an opportunity to review their attraction and recruitment strategies, not only to ensure that sufficient numbers are recruited but also to meet the ambition that forces better reflect the communities that they serve. Improvements in monitoring workforce diversity were also being led through Uplift, with national stakeholders noting the effort to standardise national data and increase declaration rates.

Barriers to diverse recruitment

To improve workforce diversity, all stakeholder interviewees described a range of barriers, typically related to attraction, which they felt needed to be addressed and overcome. These covered the broader social context, community interactions with operational policing and other workforce issues that were not related to the selection and assessment process itself:

- A lack of confidence and trust in the police, on the part of ethnic minority communities.

- The challenge faced by ethnic minority officers, due to the negative perceptions of the police in some communities.

- Limited personal familiarity with policing, such as a lack of friends or family in the service.

- Limited understanding of the officer role and policing, such as the range of career opportunities or the nature of the work.

- Policing not being viewed as a profession and whether or not new entry routes could raise the status of policing and also increase diversity.

Force strategies to improve workforce diversity

At force level, force stakeholders reported adopting strategies to reflect local demographics and experience. But there were differences in the rate of progress with implementation, as well as in the sophistication of the approaches used. Stakeholders identified a range of strategies they felt were helping to address some of the barriers and looked promising in increasing the diversity of candidates applying. However, some senior stakeholders observed that there was little evaluation of local recruitment strategies and that a better understanding of what works needed to be established.

The following attraction strategies were thought to be promising by stakeholders:

1. **Local marketing**

- Targeted to specific communities rather than blanket approaches.

- Using a range of different media, including localised letter campaigns, local radio and social media Q&A sessions.

- Using more diverse images where this had not been done previously.

2. **Engagement events, conversations and school visits**

- Tailored to different communities to explain and demystify the police officer role, career pathways and recruitment process.

- School and college leavers targeted to promote the new entry routes, particularly the Police Constable Degree Apprenticeship.

- Face to face pre-COVID-19 but forces were moving at different speeds to deliver online.

3. **A strategic focus on engaging and pursuing interested local applicants**

- To build relationships with candidates early in the process and reduce attrition prior to the sifting stage.

- Candidates provided with an opportunity to engage with the force ahead of any application, so they are more ready to apply when vacancies emerge.

4. **Involvement of staff association representatives and neighbourhood police officers**

- Staff with an understanding of, and relationship with, specific communities were important in building engagement by sharing their own stories and experience.

- Role model and buddy schemes linking ethnic minority officers with staff associations and outreach teams – to create dialogue and provide one-to-one contact through the process.

- Neighbourhood police officers used to engage communities, build trust and raise awareness of recruitment opportunities – but also important that local communities have confidence in the policing approach locally.

5. **Use of sifting prior to assessment processes**

- Sifting could increase the quality of candidates attending an assessment process and improve success rates.

- However, this needs to be accompanied by an acceptance that there will be increased attrition at the sifting stage.

6. **Local market research and analysis**

- To understand local demographics and issues and inform attraction strategies.

Support for candidates during the online process

Force support for candidates had been geared towards SEARCH® and Day One. Some forces reported having shifted to virtual support and tailored it for online assessments, while others were in the process of doing so. The pace of implementation had meant that some stakeholders felt they were unable to prepare candidates for the online assessments as they would have liked, and that the type of support they were able to provide was limited.

The following support was highlighted by stakeholders:

- maintaining contact with candidates throughout the process to keep them 'warm' for recruitment

- ensuring the College guidance around the process had been received

- chasing candidates who had not started or completed the online assessment

- delivering workshops to help improve performance, such as how to prepare for interview questions

5.3. Implementation and delivery

The overall assessment process was, on the whole, felt by force stakeholders to have run smoothly and been relatively straightforward, though a number of initial issues with delivery were identified by forces.

Timescales

Some force stakeholders felt the timescales for forces to go live with online assessments were initially unclear and subject to change. A few commented that their forces would have preferred to progress more candidates through the process, at a quicker pace. This was of particular concern where forces had a large number of candidates, requiring multiple rounds of online assessment, but were considered lower priority in the sequencing of forces for implementation. Others welcomed the data-driven and strategic approach to force prioritisation, which was viewed as placing greatest emphasis on the criticality of recruitment to the force, rather than volume.

The return of candidate results to forces was thought by many to be too slow. While some commented that the results process and supporting information sent to forces had

improved, greater automation was required. Stakeholders commented that integrating AIMS with LaunchPad, which has now happened, would help improve what was felt to be an inefficient approach to managing candidate data throughout the process. It was also suggested that this would help with the provision of more personal, detailed feedback for individual candidates, which was being requested by some force stakeholders.

Communication with candidates

Prior to undertaking the online assessment, candidates received written guidance on the process from the College. The guidance for candidates was considered by force stakeholders to be comprehensive and helpful, and updates were provided as implementation progressed and points of clarification were sought. The initial approach to communicating with candidates was felt to have possibly been confusing for some, due to the multiple emails that were sent initially by the LaunchPad system. As with the guidance, force stakeholders felt that communication with candidates improved after candidate and force feedback had been responded to by the College and its supplier.

Technology

Technical issues were raised by a small number of force stakeholders, though problems were not widespread. Interviewees reported some examples where candidates had fed back their experiences of technology failing, covering both issues with their own devices and LaunchPad connectivity. As a result, some candidates were unsure if assessments had been submitted or had to come back and undertake the assessment at a different time. The onus on the College to respond to candidate queries through the LaunchPad system was felt by some stakeholders to leave forces out of the loop and detract from the candidate experience.

Attrition

Where forces had received their assessment process results, interviewees could provide early reflections on the level of candidate attrition. Force stakeholders generally reported increased attrition compared to SEARCH® and, while some explanation was offered, further work was required to better understand the reasons why. A common perception was that it was easier for candidates to withdraw during the online process, with possibly less personal investment in the process

than with a face-to-face assessment centre that may also require an interview with the force prior to attendance. The switch from SEARCH® to online assessment was also suggested as a possible factor in some of the earliest forces to go live.

At the time of interview, some stakeholders felt that it was too early to have a sense of where attrition was most critical, the point at which candidates withdrew, or whether there was any disproportionality among underrepresented groups. Where forces had followed up with a small proportion of candidates who had withdrawn from the process, a lack of time and preparedness, or just changing their minds about pursuing policing as a career, were fed back.

College-led delivery

To establish the process as quickly as possible, delivery and administration of the online assessments for all forces was led by the College. For SEARCH® and Day One, forces could choose whether to administer their own assessment centres or send their candidates to a College-run assessment. Reflecting on the early stages of implementation, interviewees commented on preferences for future delivery models once the online assessment process had become more established.

College-led delivery was typically a preference among smaller forces and/or those that used College SEARCH® and Day One previously.

A more centralised approach was felt by force stakeholders to provide consistency of delivery, standardisation and benefits through economies of scale. One example would be a nationally managed assessor pool where assessors could be efficiently deployed to meet demand. The role of the College in quality assurance and validation was also recognised and advocated.

For smaller forces with a comparatively smaller intake of recruits to manage, the resource required to administer the assessment process at a force level was felt to be disproportionate. A College-led model was also felt by some to minimise duplicated effort across the service.

Force-led delivery

Stakeholders from forces with experience of managing their own SEARCH® and Day One assessment centres wanted

increased flexibility and control – to deliver when needed, speed up the provision of results and potentially reduce costs. The experience of delivery at the time of interview was felt by some to be prohibitive and slow, for both forces and candidates. It was felt that control of delivery would allow forces to better manage demand by processing candidates in smaller but more frequent groups, which were considered better suited to maintaining a steady flow through the pipeline.

Force stakeholders felt that the force-led model would allow for increased communication between forces and candidates, build relationships and improve support. They felt this would help to address the lack of visibility of forces in the process, one of the key limitations expressed by forces.

A blended approach was also proposed by some force stakeholders, with force-led candidate administration and College-run assessments. Such an approach was thought to increase force visibility to candidates and allow for issues to be responded to as they occurred, while maintaining consistency around assessment delivery.

5.4. Lessons learned

5.4.1. For assessment design

The move to online assessment was welcomed by stakeholders and the pandemic was felt to have been a catalyst to modernise the selection and assessment process, removing any previous blockages to change. Among all stakeholders there was no strong desire to return to SEARCH® or Day One assessment centres, when COVID-19 restrictions allowed. But there was support for an assessment model that incorporated more face-to-face and interactive aspects.

Face-to-face interviews, either in person or online, were proposed as it was felt this would allow forces to meet and better assess candidates prior to appointment. Several stakeholders felt this more personable approach would address the key limitation of the recorded interview format used within the online assessment, which was the lack of two-way communication.

"But I do also think it's important when you're hiring someone that you're hiring them for the values of your organisation, for your key mission and objectives, and therefore an option, or manifestation of a more personalised process that gives the candidate and the assessor the opportunity to check suitability, both for the employer and the employee, is an essential part of selection."

[Stakeholder interview]

In much the same way, some stakeholders also felt that a more interactive approach to the briefing exercise would allow for probing questions that may improve candidate performance. The written exercise was viewed to have worked well, with the structured format being well suited to the online process.

The high value that stakeholders placed on interaction was also expressed in support of the role-play exercises undertaken at SEARCH® and Day One. Role play was viewed as an important element absent from the online

assessment, one that provided candidates with the closest representation of the job by placing them in a challenging encounter.

While the role plays at SEARCH® and Day One employed actors to perform different scenarios, which candidates were required to respond to, a few interviewees suggested that the functionality of LaunchPad (or other platforms that can introduce gamification) could be explored in the future. Live video, virtual reality, digital role play or simulation could allow for new exercises that can assess quick decision making.

A small number of interviewees were more cautious about including a face-to-face interview within the online assessment process due to concerns over whether this would introduce unconscious bias and increase adverse impact. The emerging assessment process results (viewed by stakeholders at the time of interview) provided an indication that adverse impact on ethnic minorities was improving. Reflecting on this, and the initial positive reception of the online process, a few interviewees questioned whether additional exercises were indeed required, on the assumption

that the process had recruited candidates that were fit for purpose.

5.4.2. For project management and delivery

College resourcing

College staff were commended by stakeholders for their technical expertise and the strong relationships they had quickly established with forces. To aid delivery, single points of contact were put in place to connect the College with force recruitment leads. Interviewees welcomed the responsiveness of the College, both in the way urgent issues were typically dealt with quickly and cooperatively and in the provision of assessment outcomes data for forces, which was felt to be timely and useful. The College's expertise in the analysis of outcome data was regarded as highly valuable to the service and opportunities to share knowledge further would be welcomed.

Some senior stakeholders felt that College resources had been stretched over capacity and concerns were raised over the sustainability of the current approach. Having worked to deliver the new process, a few stakeholders also felt that

expectations on what could be delivered with the available resources may need to be reset.

Project management

While the pace and impact of implementation was highlighted by interviewees overall, a few provided feedback to help inform ongoing and future project management. It was suggested by some stakeholders that the focus on speed meant that robust project management practices were not fully established. Milestones, deliverables, outcomes, governance arrangements and costs would have benefited from further clarity and development. Additional delivery, finance and project management expertise were suggested as being helpful to the future implementation of such a large-scale project.

Cost to forces

A few force stakeholders commented that the pricing structure, and how this was initially communicated to forces, required more clarity.

A small number of interviewees felt that the College was slow to understand the contractual obligation faced by some forces

with universities delivering Police Constable Degree Apprenticeship places. These forces needed to go live at a certain time and achieve minimum numbers for the university course otherwise there would be penalties.

Future models

At the time of interview, force stakeholders suggested it was too early for forces to make any decisions around future delivery models. To make an informed decision, further information was required on the costs, delivery, understanding adverse impact and diversity outcomes.

Forces used to managing their own assessment centres felt that moving to a force-led model would provide increased control and flexibility. Reflecting on future delivery models, a few senior stakeholders were wary of different assessment approaches being adopted in different forces, if there was too much flexibility for forces. They added the consistency of assessment introduced through the online process was to be encouraged and that the College should set standards across all areas of selection, covering roles other than police constable.

6. Conclusion and next steps

Closing the gap in ethnic minority performance

Following the introduction of the online assessment process, more candidates from ethnic minority backgrounds have been successful, compared to SEARCH®. This is evident in an improved AIR of 85.1%, which is now above the 80% threshold for determining adverse impact in selection processes.

There were also no statistically significant differences in the views of White and ethnic minority candidates regarding their satisfaction with the online assessment process and the way they were able to demonstrate their abilities.

However, there is still a small but statistically significant gap between the results of White candidates and those from ethnic minority backgrounds across all exercises. Some ethnic minority groups are also still experiencing an adverse impact, which suggests that further efforts are required to make the policing workforce more representative.

Forces have shared examples of practice deemed to be effective in attracting candidates from ethnic minority backgrounds, which may assist others in identifying potential attraction strategies that support diverse recruitment.

The review by occupational psychologists at Fairway concluded that the process was fair, and the immediate recommendations implemented. But they added further research and evaluation was required to better understand and explain the differences in success rates between candidates from White and ethnic minority backgrounds, as well as what attraction strategies were most effective.

The College will undertake further analysis to better understand the differences between groups, including work to understand whether other characteristics, such as sex or age, moderate the outcomes for ethnicity. The integration of AIMS will also allow for additional factors to be included within future analysis, such as language proficiency, educational background and socioeconomic status.

The College will also commission a qualitative review to understand the gap in outcomes between candidates from different ethnic groups, and identify actions that could be taken to reduce them.

Assessor training

Findings from the survey and IRR study showed that assessors were less confident and less consistent in their scoring of the competency-based interview, and would value more support and training in the new online model.

After receiving this and ongoing feedback from assessors, a number of enhancements have been made to enhance the quality and reliability of assessors. Training and support for assessors has been developed to provide:

- specific skills development for understanding biases and barriers to effective assessment, scoring scales, competencies and exercises, and using LaunchPad
- more detailed feedback on assessor scoring
- opportunities for refresher training
- additional examples to support better benchmarking
- targeted training for underperforming assessors
- further guidance on common queries and issues

The IRR findings can also be used as a baseline for further analysis for continuous improvement purposes and to inform assessor training and quality assurance processes.

Assessed exercises

The online assessment process was developed as an emergency interim measure to support policing to continue recruitment during the pandemic. There was consensus among stakeholders that the online assessment process met this objective and that forces were satisfied that the process was fit for purpose.

Candidates also expressed high levels of overall satisfaction that were consistent with SEARCH and Day One®.

However, several interviewees expressed support for exercises that would provide opportunity for further interaction and assess real-time decision making. They added this may also go some way to supporting more candidates feel that they can demonstrate their full range of abilities.

The College will continue to review the assessment exercises and assessment model, and identify ways of improving the reliability and validity of assessment. Any changes will be incremental, with ongoing reviews of the impact on diversity outcomes.

Any adjustment to the exercises will need to be monitored for any potential changes in the AIR.

Future delivery

Force perspectives on future delivery models were shaped by the size of the force, the demand for recruitment and whether College SEARCH® or Day One assessment centres were used previously. Forces used to managing their own assessment centres felt that moving to a force-led model would provide increased control and flexibility, and allow for more direct communication with candidates.

The College of Policing are committing to delivering initial recruit assessments in their current online format, subject to further enhancements, until 31st March 2022, with a review point in January 2022.

Wider recruitment process improvements

A number of other improvements to the wider recruitment process have also been introduced to help increase workforce diversity.

- The National Police Constable Application Form has now been revised and updated following a comprehensive review. All forces now have access to the form, which has been designed to be welcoming, use inclusive language and support successful completion by applicants with diverse characteristics.

- New guidance and tools have been developed for recruitment teams on candidate management throughout the recruitment process. This includes a tool to calculate and interpret adverse impact (disparity in pass rates) plus advice on what factors to address to remove any adverse impact.

- Local selection process workshops have been delivered to forces and national stakeholders to focus on local recruitment processes. Specifically they focus on understanding and removing adverse impact for underrepresented groups within pre-sift processes and later local interviews. These teachable sessions upskilled force recruitment teams and shared the College subject-matter experts' knowledge.

- The College is now scoping a national standard for pre-sift selection processes and local interviews, which are managed within forces – to increase consistency and remove adverse impact.

Annex A. Methods

A1. Evaluation team

The evaluation has been undertaken by an in-house College team of researchers and statisticians, independent from the Selection and Assessment team responsible for implementing the online assessment process with forces. College research staff are members of the Government Social Research (GSR) profession. As such, they are bound by the Civil Service Code (and its core values of integrity, honesty, objectivity and impartiality), as well as the professional standards set out in the specific GSR code, which have been designed to ensure quality, rigorous social research and analysis.

The online assessment outcome data has been analysed by chartered occupational psychologists in the Selection and Assessment team as part of the ongoing monitoring and reporting of results to forces and other stakeholders.

The report has been reviewed by an independent academic peer reviewer. The IRR study was also subject to academic peer review at both the design and reporting stages.

A2. IRR study

The process below was carried out for each of the three exercises being assessed. At a late stage, the design was altered from 200 cases examined by two assessors (the original score plus one re-mark) to 100 cases that would be marked three times (two re-marks plus the original score). This change was made when we became aware that there was a risk that the original scores would not be readily available.

1. Of the available cases in the database that had already been marked, a sample of 100 cases was randomly selected. This was a compromise between the number that was practical to re-mark, and the precision of the final results.

2. For each exercise, a pool of 10 randomly selected assessors was created from the ~450 examiners in the database. As some examiners assess more cases than others, we used a stratified random selection with the intention that this would result in a better 'real world' estimate of the IRR. The examiners were ranked by the number of times they had been employed and then split

into ten 'strata', each representing approximately 10% of the total. This resulted in a small group of high frequency examiners at one extreme, and a large group of lesser-used examiners at the other. One examiner was randomly selected from each stratum and, if they were unavailable, an alternative was randomly selected from the same group.

3. The cases and examiners were randomly allocated, with a slight adjustment in the final few allocations to ensure that each examiner marked 20 cases.

4. The links to the exercises and communications with the examiners were handled by the College administrative team. The marks were returned using an Excel spreadsheet (rather than being entered into the online database). The examiners were aware that they were part of a study and, although some were unavailable in the short window we gave them for marking, none of those selected declined to take part.

5. The assessors for both the original marking and the re-marking were from the same 'pool' and the conditions for the marking were the same with the exception that the scores for the re-marking were entered into a spreadsheet rather than an online form.

Key limitations

The study was not conducted blind and the assessors who conducted the re-marking knew they were participating in a research study. This may have resulted in a trial or Hawthorne effect whereby research participants adjust their behaviour when they know they are being observed for a study.

Due to sample sizes, it has not been possible to analyse findings by ethnicity.

Annex B. Online assessment process outcomes

B1. Outcome data tables for protected characteristics (14 January 2021)

Table 1. Summary results

Summary results	Online assessment process				SEARCH®	
	Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Total no.	4,756	12,450	17,206		47,507	
Percentage	27.64	72.36			71.30%	

Table 2. Ethnicity 2+1

Ethnicity 2+1		Online assessment process				SEARCH®	
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Not stated	Total no.	251	674			12,733	
	Percentage	27.14%	72.86%	925		71.60%	
White	Total no.	3,812	10,614		Majority group	31,408	
	Percentage	26.42%	73.58%	14,426		73.40%	
Ethnic minority	Total no.	693	1,162			3,365	
	Percentage	37.36%	62.64%	1,855	85.14%	55.80%	76.02%

Table 3. Ethnicity 5+1

Ethnicity 5+1		Online assessment process				SEARCH®	
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Not stated	Total no.	251	674			12,733	
	Percentage	27.14%	72.86%	925		71.60%	
White	Total no.	3,812	10,614			31,408	
	Percentage	26.42%	73.58%	14,426	Majority group	73.40%	
Mixed/multiple ethnic groups	Total no.	132	374			950	
	Percentage	26.09%	73.91%	506	100.46%	70.60%	96.19%
Asian/Asian British	Total no.	427	588			1,631	
	Percentage	42.07%	57.93%	1,015	78.74%	51.90%	70.70%
Black/African	Total no.	90	125			487	
	Percentage	41.86%	58.14%	215	79.02%	49.20%	67.03%
Other ethnic group	Total no.	28	54			297	
	Percentage	34.15%	65.85%	82	89.50%	53.40%	72.75%
Ethnic minority – not collected	Total no.	16	21				
	Percentage	43.24%	56.76%	37	N/A	n/a	n/a

Table 4. Ethnicity 18+1

Ethnicity 18+1		Online assessment process			
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio
Prefer not to disclose	Total no.	251	674		
	Percentage	27.14%	72.86%	925	N/A
White – English/Welsh/Scottish/Northern Irish/British	Total no.	3,154	9,360		
	Percentage	25.20%	74.80%	12,514	Majority group
White – Irish	Total no.	17	41		
	Percentage	29.31%	70.69%	58	94.51%
White – Gypsy	Total no.	27	88		
	Percentage	23.48%	76.52%	115	102.31%
White – any other White	Total no.	274	390		
	Percentage	41.27%	58.73%	664	78.53%
Mixed – White and Black Caribbean	Total no.	50	125		
	Percentage	28.57%	71.43%	175	95.50%
Mixed – White and Black African	Total no.	15	29		
	Percentage	34.09%	65.91%	44	88.12%
Mixed – White and Asian	Total no.	33	114		
	Percentage	22.45%	77.55%	147	103.68%
Mixed – Any other mixed/multiple ethnic background	Total no.	28	95		
	Percentage	22.76%	77.24%	123	103.26%
Asian – Indian	Total no.	95	181		
	Percentage	34.42%	65.58%	276	87.68%
Asian – Bangladeshi	Total no.	55	54		
	Percentage	50.46%	49.54%	109	66.24%

Ethnicity 18+1		Online assessment process			
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio
Asian – Chinese	Total no.	16	15		
	Percentage	51.61%	48.39%	31	64.69%
Asian – Any other Asian	Total no.	43	46		
	Percentage	48.31%	51.69%	89	69.10%
Black – African	Total no.	57	53		
	Percentage	51.82%	48.18%	110	64.42%
Black – Caribbean	Total no.	25	52		
	Percentage	32.47%	67.53%	77	90.29%
Black – Any other Black African/Caribbean background	Total no.	7	17		
	Percentage	29.17%	70.83%	24	94.70%
Any other ethnic group – Arab	Total no.	6	9		
	Percentage	40.00%	60.00%	15	80.22%
Any other ethnic group	Total no.	23	45		
	Percentage	33.82%	66.18%	68	88.48%

Table 5. Sex

Sex		Online assessment process				SEARCH®	
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Not stated	Total no.	374	1,023			0	
	Percentage	26.77%	73.23%	1,397	N/A	0.00%	
Female	Total no.	1,682	5,128			15,993	
	Percentage	24.70%	75.30%	6,810	107.59%	75.74%	109.38%
Male	Total no.	2,699	6,295			31,514	
	Percentage	30.01%	69.99%	8,994	Majority group	69.24%	
Trans[7] (legacy data)	Total no.	1	4			0	
	Percentage	20.00%	80.00%	5	114.30%	0.00%	n/a

---

[7] Category changed to 'Intersex' for future data collection and reporting.

Table 6. Age

Age		Online assessment process				SEARCH®	
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Not stated	Total no.	223	540			0	
	Percentage	29.23%	70.77%	763	N/A	0.00%	
Under 26	Total no.	2,623	6,622			5,252	
	Percentage	28.37%	71.63%	9,245	96.78%	60.94%	83.95%
26-40	Total no.	1,648	4,693			19,991	
	Percentage	25.99%	74.01%	6,341	Majority group	72.59%	
41-55	Total no.	256	580			12,260	
	Percentage	30.62%	69.38%	836	93.74%	75.82%	104.46%
Over 55	Total no.	6	15			4,738	
	Percentage	28.57%	71.43%	21	96.51%	73.95%	101.88%

Table 7. Disability

Disability							
		Online assessment process				**SEARCH®**	
		Unsuccessful candidates	**Successful candidates**	**Total number of candidates**	**Adverse impact ratio**	**SEARCH® successful candidates**	**Adverse impact ratio**
Not stated	Total no.	542	1,671			0	
	Percentage	24.49%	75.51%	2213	N/A	0.00%	
Disability	Total no.	233	548			1,489	
	Percentage	29.83%	70.17%	781	97.47%	69.13%	96.91%
No disability	Total no.	3,981	10,231			44,374	
	Percentage	28.01%	71.99%	14,212	Majority group	71.33%	

Table 8. Religion

Religion		Online assessment process				SEARCH®	
		Unsuccessful candidates	Successful candidates	Total number of candidates	Adverse impact ratio	SEARCH® successful candidates	Adverse impact ratio
Not stated	Total no.	1,107	2,968			12,829	
	Percentage	27.17%	72.83%	4,075	N/A	72.02%	
Christian	Total no.	1,091	2,806		Majority group	12,366	
	Percentage	28.00%	72.00%	3,897		70.78%	
Jewish	Total no.	2	23			94	
	Percentage	8.00%	92.00%	25	127.77%	75.20%	106.24%
Muslim	Total no.	329	410			1,453	
	Percentage	44.52%	55.48%	739	77.05%	49.35%	69.73%
No religion or belief	Total no.	2,051	5,793			16,574	
	Percentage	26.15%	73.85%	7,844	102.57%	74.22%	104.86%
Sikh	Total no.	42	83			253	
	Percentage	33.60%	66.40%	125	92.22%	60.38%	85.30%
Other	Total no.	84	290			214	
	Percentage	22.46%	77.54%	374	107.69%	62.21%	87.89%
Hindu	Total no.	33	42			176	
	Percentage	44.00%	56.00%	75	77.77%	53.66%	75.81%
Buddhist	Total no.	17	35			108	
	Percentage	32.69%	67.31%	52	93.48%	69.68%	98.44%

B2. Statistical analysis of online assessment process outcomes

Online assessment process exercises – summary

- SJT: The 4,011 White candidates compared to 499 candidates from a Minority Ethnic background demonstrated significantly higher mean exercise scores, t = -3.920, p = .001.

- CBI: The 10,560 White candidates compared to 1,318 candidates from a Minority Ethnic background demonstrated significantly higher mean exercise scores, t = -4.63, p = .001.

- Briefing: The 10,502 White candidates compared to 1,303 candidates from a Minority Ethnic background demonstrated significantly higher mean exercise scores, t = -10.32, p = .001.

- Written: The 10,645 White candidates compared to 1,324 candidates from a Minority Ethnic background demonstrated significantly higher mean exercise scores, t = -8.782, p = .001.

These are statistically significant, however, the difference in the effect sizes are small/small-medium and smaller than at SEARCH® (see tables below).

Table 1. Ethnicity analysis of individual exercises at the online assessment process (t-test)

Exercise	Ethnic minority			White				
	No.	Mean	SD	No.	Mean	SD	t-test	Sig
SJT	499	86.21	5.74	4,011	87.28	5.83	-3.92	0.001
CBI	1,318	61.28	15.6	10,560	63.39	15.6	-4.63	0.001
Briefing	1,303	57.79	14.5	10,502	62.19	14.64	-10.32	0.001
Written	1,324	55.77	16.78	10,645	60.05	16.27	-8.782	0.001

Table 2. Ethnicity analysis of individual exercises at the online assessment process (Cohen's d)

Exercises	White N	Ethnic minority N	White mean	Ethnic minority mean	SD_White	SD_Ethnic minority	Diff	Pooled_SD	Cohen's_d	Effect size
SJT	4,011	499	87.3	86.2	5.8	5.7	1.1	5.79	0.18	Small
CBI	10,560	1,318	63.4	61.3	15.6	15.6	2.1	15.61	0.14	Small
Briefing	10,502	1,303	62.2	57.8	14.6	14.5	4.4	14.57	0.30	Small-medium
Written	10,645	1,324	60.1	55.8	16.3	16.8	4.3	16.53	0.26	Small-medium

Table 3. Ethnicity analysis of individual exercises at SEARCH® (Cohen's d)

Exercises	White N	Ethnic minority N	White mean	Ethnic minority mean	SD_White	SD_Ethnic minority	Diff	Pooled_SD	Cohen's_d	Effect Size
Zajak (Interactive Exercise)	42799	6032	55.0359	49.6620	11.26819	12.63453	5.4	11.97	0.45	Medium
Swift (Interactive Exercise)	42799	6032	54.8713	49.8094	11.33574	12.16638	5.1	11.76	0.43	Medium
Kowalski (interactive Exercise)	42799	6032	56.8092	53.6279	10.26178	11.60731	3.2	10.96	0.29	Small to medium
Garner (interactive Exercise)	42799	6032	54.9152	51.8457	10.57343	11.48307	3.1	11.04	0.28	Small to medium
Interview	42799	6032	55.1285	53.8032	13.05993	13.45034	1.3	13.26	0.10	Small
Verbal Ability Test	42799	6032	70.0502	58.6094	13.30906	16.26555	11.4	14.86	0.77	Medium to Large
Numerical Ability Test	42799	6032	63.2627	52.2341	19.52389	20.29351	11.0	19.91	0.55	Medium
Written	37619	3059	69.3038	60.2556	10.61418	13.62645	9.0	12.21	0.74	Medium to Large

Annex C. Candidate and assessor survey results

C1. Candidate survey respondent demographics

Table 1. Age

	Percentage of respondents (%)
	n=931
Under 26	57
26-40	38
41-55	6
Over 55	n<5
Prefer not to say	n<5

Table 2. Disability

	Percentage of respondents (%)
	n=931
Disabled	7
Not disabled	93

Table 3. Ethnicity

	Percentage of respondents (%)
	n=931
White	92
Ethnic minority	7
Prefer not to say	n<5

Table 4. Sex

	Percentage of respondents (%)
	n=931
Female	41
Male	59
Prefer not to say	n<5

Table 5. Parent/guardian highest level qualification

	Percentage of respondents (%)
	n=931
At least one has a degree level qualification	29
Qualifications below degree level	56
No formal qualifications	7
Don't know	6
Not applicable	1
Prefer not to say	1
Other	1

C2. Statistical analysis of differences in candidate satisfaction between ethnic groups

Satisfaction with overall delivery of the online assessment process

How satisfied were you with the overall delivery of the online assessment process?

		Very satisfied	Satisfied	Neutral	Dissatisfied	Very dissatisfied	Total
Ethnicity	White	392	332	77	43	13	857
	Ethnic minority	30	23	7	7	0	67
Total		422	355	84	50	13	924

Ranks

Ethnicity	N	Mean rank	Sum of ranks
White	857	461.22	395263.00
Ethnic minority	67	478.91	32087.00
Total	924		

Test statistics[a]

	Q31
Mann-Whitney U	27610.000
Wilcoxon W	395263.000
Z	-.568
Asymp. Sig. (2-tailed)	.570

a. Grouping variable: Ethnic_Group

Demonstration of abilities

Please indicate your level of agreement with the following statement: 'The online assessment process allowed me to demonstrate my full range of skills and abilities'

		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	
Ethnicity	White	142	324	216	128	47	857
	Ethnic minority	16	26	14	6	5	67
Total		158	350	230	134	52	924

Ranks

Ethnicity	N	Mean rank	Sum of ranks
White	857	465.99	399352.50
Ethnic minority	67	417.87	27997.50
Total	924		

Test statistics[a]

	Q32
Mann-Whitney U	25719.500
Wilcoxon W	27997.500
Z	-1.480
Asymp. Sig. (2-tailed)	.139

a. Grouping variable: Ethnic_Group

C3. Assessor survey respondent demographics

Table 1. Age

	Percentage of respondents (%)
	n=253
Under 26	n<5
26-40	5
41-55	38
Over 55	55
Prefer not to say	n<5

Table 2. Ethnicity

	Percentage of respondents (%)
	n=253
White	88
Ethnic minority	10
Prefer not to say	2

Table 3. Sex

|	Percentage of respondents (%)
	n=253
Female	50
Male	49
Prefer not to say	n<5

Annex D. IRR analysis

D1. Results for each exercise

Competency-based interview

There were 279 marks used for the analysis across 93 cases (three per case). The average (mean) score was 2.81 with a standard deviation (SD) of 0.88, and a pass rate of 63% (grades of three and above). The distribution of grades for the competency-based interview (2a) exercise is shown below.

CBI (2a) - Distribution of Grades (n=279)
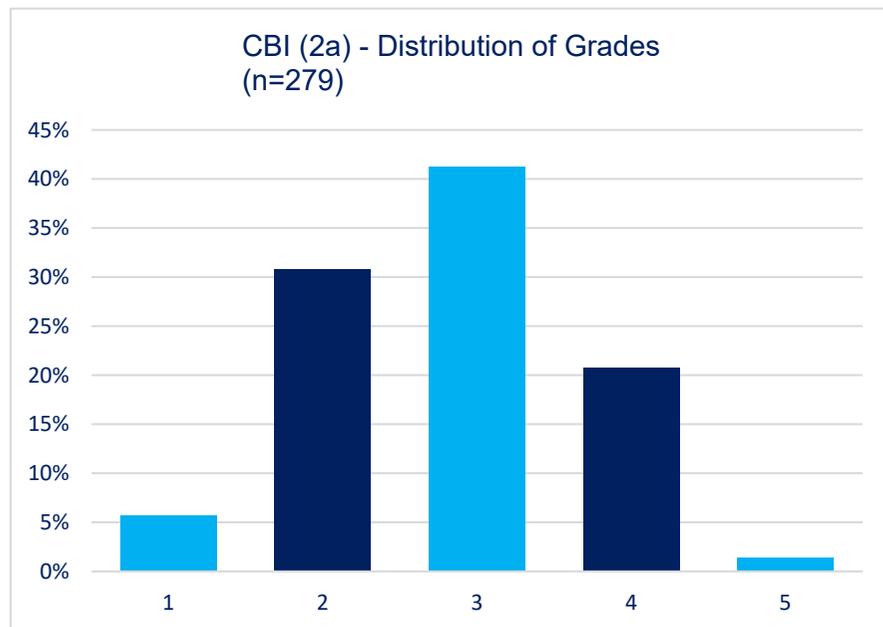
In comparing the grades given by the new and original assessors for this exercise, there is a noticeable difference in the distribution of scores, with the new markers (ie, those recruited for the study) tending to score applicants lower than they were originally marked (below).

CBI (2a) - Proportion of Grades Given By New and Original Assessors
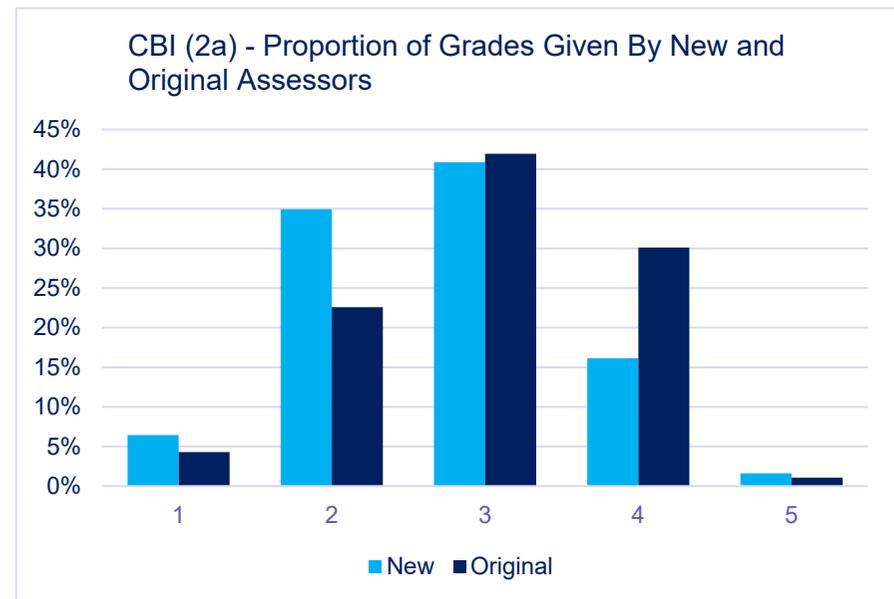
On closer examination, the scores given by two to three of the new markers appeared to be noticeably adrift of their colleagues. Further investigation provided evidence that this was primarily due to two of the 10 new assessors providing scores that were much lower than their colleagues. The

difference in mean scores for each of these two assessors, when compared to all other scores, was statistically significant (two independent sample t-test, unequal sample sizes and variance, $p<0.01$ and $p<0.05$). The results from a non-parametric Mann-Whitney were similarly significant.

This skew in marking is possibly caused by the chance, random nature of the assessor selection or a Hawthorne or trial effect where subjects adjust their behaviour when they know they are being observed or are the subject of a study.

This is noticeable when comparing the ICC for the total sample for this exercise (original scores plus new marks) compared to the sample of just the new re-markers.

Intraclass correlation coefficient (total sample)

| | Intraclass correlation | 95% confidence interval | | F test with true value 0 | | | |
		L.Bound	U.Bound	Value	df1	df2	Sig
Single measures	.290	.161	.424	2.226	92	186	.000
Average measures	.551	.366	.688	2.226	92	186	.000

One-way random effects model where people effects are random.

Intraclass correlation coefficient (new marks only)

| | Intraclass correlation | 95% Confidence Interval | | F test with true value 0 | | | |
		L.Bound	U.Bound	Value	df1	df2	Sig
Single measures	.240	.040	.422	1.632	92	93	.010
Average measures	.387	.077	.594	1.632	92	93	.010

One-way random effects model where people effects are random.

Unusually, the ICC for the re-markers alone (0.24) is lower than that of the re-markers and the original scores combined (0.29), implying that there is a substantial source of variation within the new assessors' scores.

We have reported the combined score, as this is (a) from a larger sample, and (b) likely to be closer to the true underlying rate.

Written exercise

There was no evidence of a difference in the scores of the new (IRR study) and original assessors for the written exercise.

There were 300 marks used for the analysis across 100 cases. These consisted of 200 from this exercise and 100 from the original scores. The average (mean) score was 3.03 with a standard deviation (SD) of 0.86, and a 76% pass rate (scores of 3 and above). The distribution of grades for the written exercise (3a) is shown below.
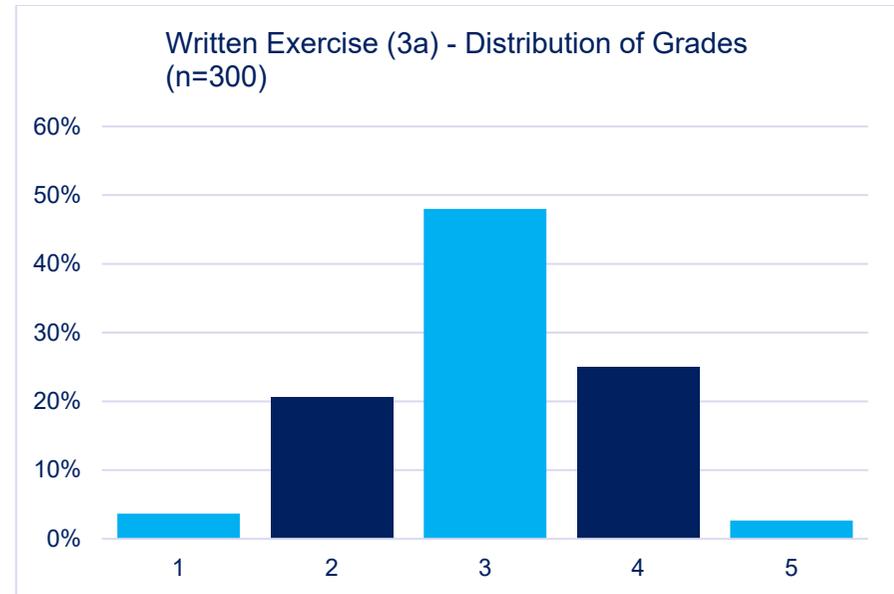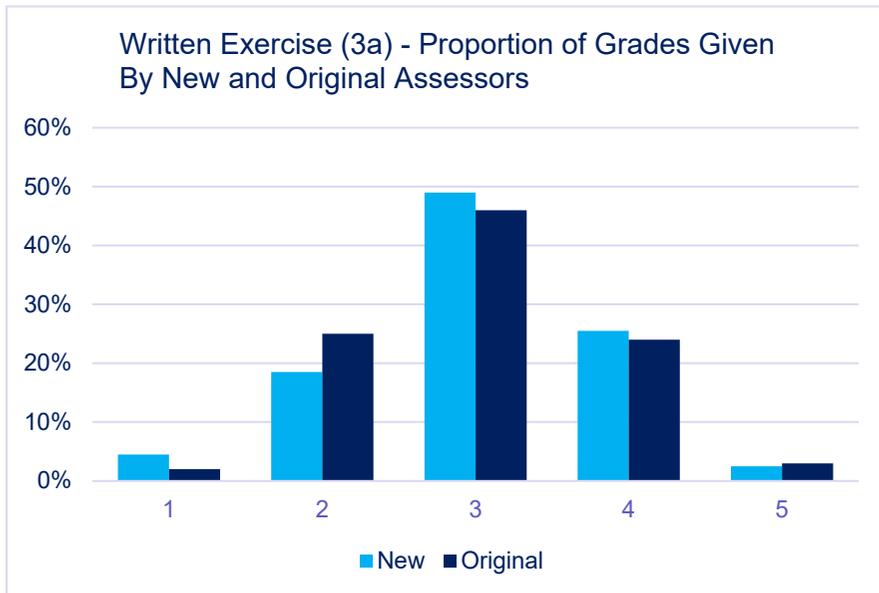
Written Exercise (3a) - Distribution of Grades (n=300)
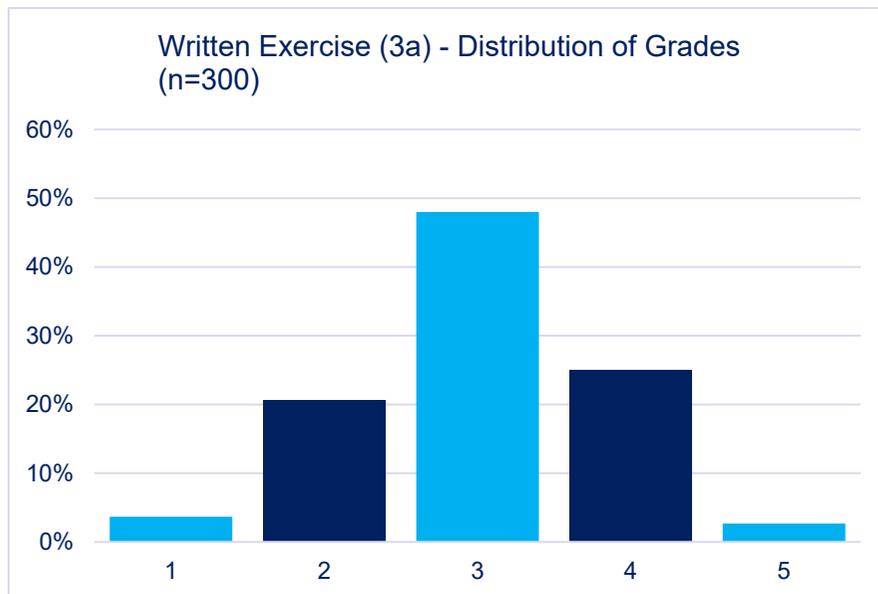
The ICC for the assessment is shown below.

Intraclass correlation coefficient (written)

	Intraclass correlation	95% confidence interval		F test with true value 0			
		L.Bound	U.Bound	Value	df1	df2	Sig
Single measures	.538	.426	.643	4.495	99	200	.000
Average measures	.778	.690	.844	4.495	99	200	.000

One-way random effects model where people effects are random.

The single measures figure (0.54) is the ICC for the current situation where we have one rater per exercise. The average measures figure (0.78) is the value that would be achieved if we used multiple markers for each exercise and averaged the scores. The average value, by its very nature, will always be higher than the single measures score, but implementing such a scheme will be impractical in many circumstances due to the relatively low increase in reliability for such a large increase in resources and cost.

Briefing exercise

There were 276 marks used for the analysis across 92 cases. The average (mean) score was 3.12 with an SD of 0.83, and a 77% pass rate (scores of 3 and above). A comparison of the grades given by the new and original scores for the briefing exercise is shown below.

Written Exercise (3a) - Proportion of Grades Given By New and Original Assessors

Written Exercise (3a) - Distribution of Grades (n=300)

The difference between the original and new assessors was not statistically significant (two independent sample t-test, p=0.11).

The SPSS output for the ICC calculation is shown below.

Intraclass correlation coefficient (briefing)

	Intraclass correlation	95% confidence interval		F test with true value 0			
		L.Bound	U.Bound	Value	df1	df2	Sig
Single measures	.340	.211	.472	2.547	91	184	.000
Average measures	.607	.445	.728	2.547	91	184	.000

One-way random effects model where people effects are random.

D2. Alternative scoring methods

Alongside the overall mark for each exercise the assessors also look for and score four to five desirable behaviours, for example, 'Integrity' and 'We take ownership'. Each behaviour consists of another four to five sub-behaviours so this is potentially a richer dataset than the single overall mark given by the assessors. The analysis in this section explores whether these scores are a more consistent marking scheme than the single overall mark.

Method

For each of the exercises the overall mark for the four to five behaviours were averaged for each assessment. This produced a score from each of the three assessors similar to the overall mark, but on a near continuous scale. This single average of the behaviours was then used in place of the overall mark and the IRR calculated using the same ICC method as before.

Results

The results of the analysis alongside the original IRR scores are shown below.

Exercise	Original IRR	Alternative IRR	Lower CI	Upper-CI
Written	0.54	0.57	0.46	0.67
Briefing	0.34	0.4	0.28	0.53
Competency-based interview	0.29	0.34	0.21	0.47

In each case there has been a moderate or marginal improvement in the IRR. Both the briefing and competency-based interview exercises increase their score by around 0.05 to 0.06 points, which is an increase of ~18% on the old scores. It should be kept in mind that some of this increase may be down to the increased granularity of the averaged scores. Under this scheme, the briefing exercise would go from 'fair' to borderline 'moderate'.

An alternative way of comparing the data is to look at the average difference in grades between any two assessors marking the same candidate. These are shown in the table below.

	Original	Alternative
Written	0.55	0.55
Briefing	0.66	0.64
Competency-based interview	0.77	0.70

For the written exercise there is no overall difference between the two scoring methods. In both cases the average gap between re-marked papers was just over half a grade. This could be interpreted as demonstrating a good match between the behaviour scores and the overall score given by an assessor. The biggest improvement is in the competency-based interview exercise where the average gap fell from 0.77 to 0.70 grades.

One other line of analysis were explored. We looked at whether a combination of the behaviour scores and the overall mark produced a better IRR score than the behaviour scores alone. However, the change was negligible and did not affect the IRR scores to two decimal places.

About the College

We're the professional body for the police service in England and Wales.

Working together with everyone in policing, we share the skills and knowledge officers and staff need to prevent crime and keep people safe.

We set the standards in policing to build and preserve public trust and we help those in policing develop the expertise needed to meet the demands of today and prepare for the challenges of the future.

college.police.uk