



National Audit Office

Report

by the Comptroller
and Auditor General

Home Office

Investigation into the response to cheating in English language tests

What this investigation is about

1 In 2014 the BBC's *Panorama* programme uncovered examples of organised fraud taking place during Secure English Language Tests (SELTs) required under immigration rules for students and other visas. The Home Office (the Department) responded by suspending the licences of test centres and revoking the visas of people accused of cheating. A number of people have protested their innocence. The Home Affairs Select Committee has carried out an inquiry, and the House of Commons has debated the matter, without achieving a clear resolution.

2 The Department works with a number of different bodies regarding SELTs. This includes:

- SELT providers, who develop, administer and mark the tests;
- student sponsors such as private colleges, who must have a licence with the Department to admit students with General student visas (Tier 4) and must remain compliant with the Department's terms; and
- visa applicants, including students who apply to study with Tier 4 sponsors and must demonstrate their English language proficiency before doing so, in some cases by passing a SELT.

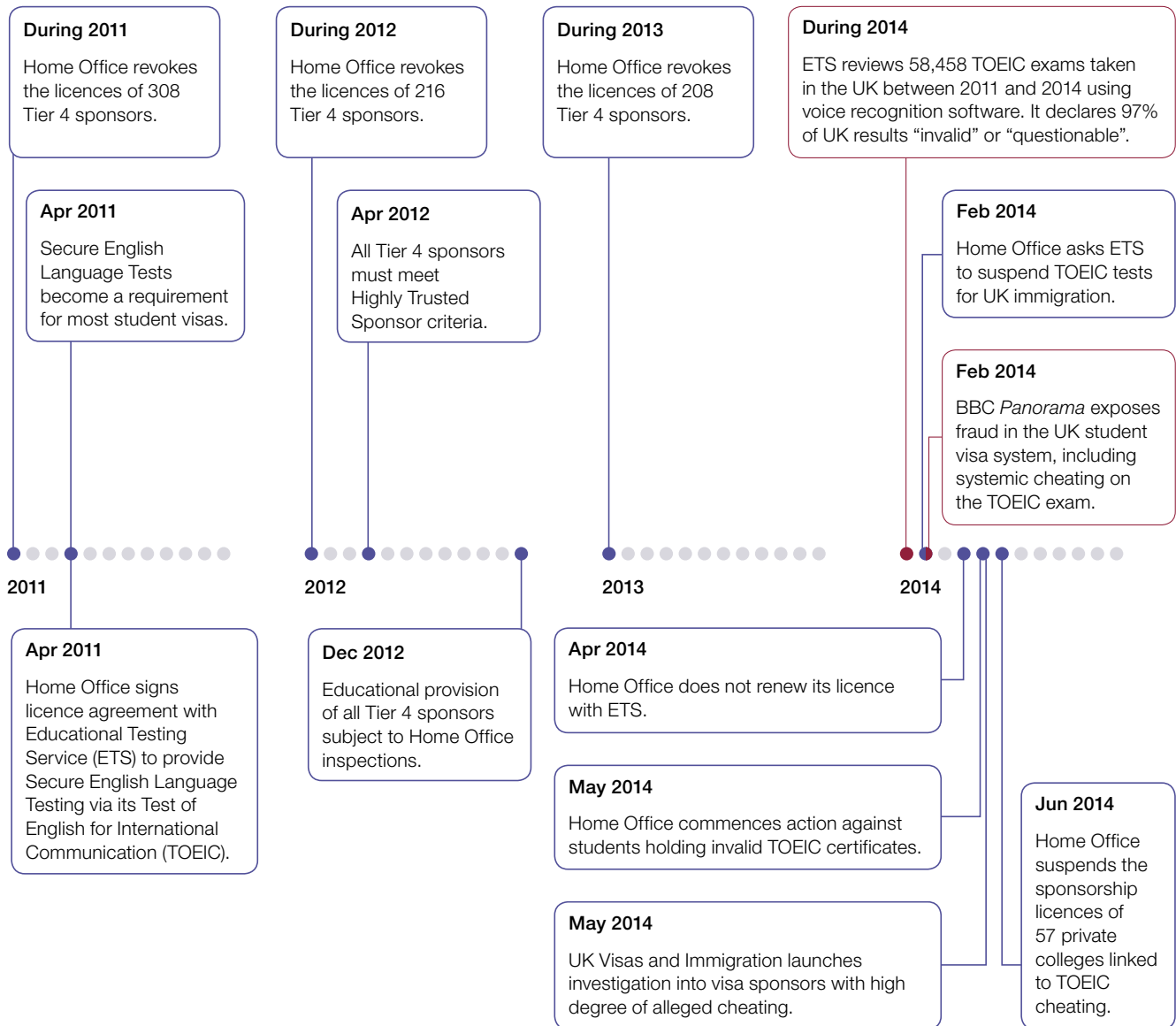
3 This investigation examines the Department's response to suspected cheating in the Test of English for International Communication (TOEIC), a SELT administered by Educational Testing Service Global BV, a subsidiary of Educational Testing Service USA (ETS), a non-profit educational testing and assessment organisation. We consider how the Department assessed the number of people who may have cheated and what action it took.

4 We looked at the Department's approach to identifying cheats and at its quality assurance work. In particular, we looked at its estimates of the number of people cheating, the steps it took to assess the evidence, and what happened to people subsequently. We examined the Department's data and documents, reviewed published material and met with a number of stakeholders. **Figure 1** on pages 6 and 7 sets out the chronology of major events. **Figure 2** on pages 8 and 9 sets out what happened to the people affected. Appendix One sets out our methodology.

5 A number of factors make it challenging to provide a consistent picture over time about what happened to the individuals concerned. These factors include: the different kinds of monitoring the Department undertook between 2014 and the present, the evolution of the Department's approach to legal cases and changes in individuals' circumstances which have in some cases altered their immigration status.

Figure 1
Timeline of events

Home Office action against fraud in English language tests including the Test of English for International Communication (TOEIC) spans many years



□ Public events □ Home Office actions

Notes

- From 6 April 2015 only human rights, protection claims, or European Economic Area decisions would provide foreign students with a right of appeal.
- People can apply for a Tier 4 (General) student visa to study in the UK if they are 16 and over and have been: offered a place on a course; can speak, read, write and understand English to an accepted standard; have enough money to support themselves and pay for their course; are from a country that is not in the European Economic Area (EEA) or Switzerland; and meet other eligibility requirements.

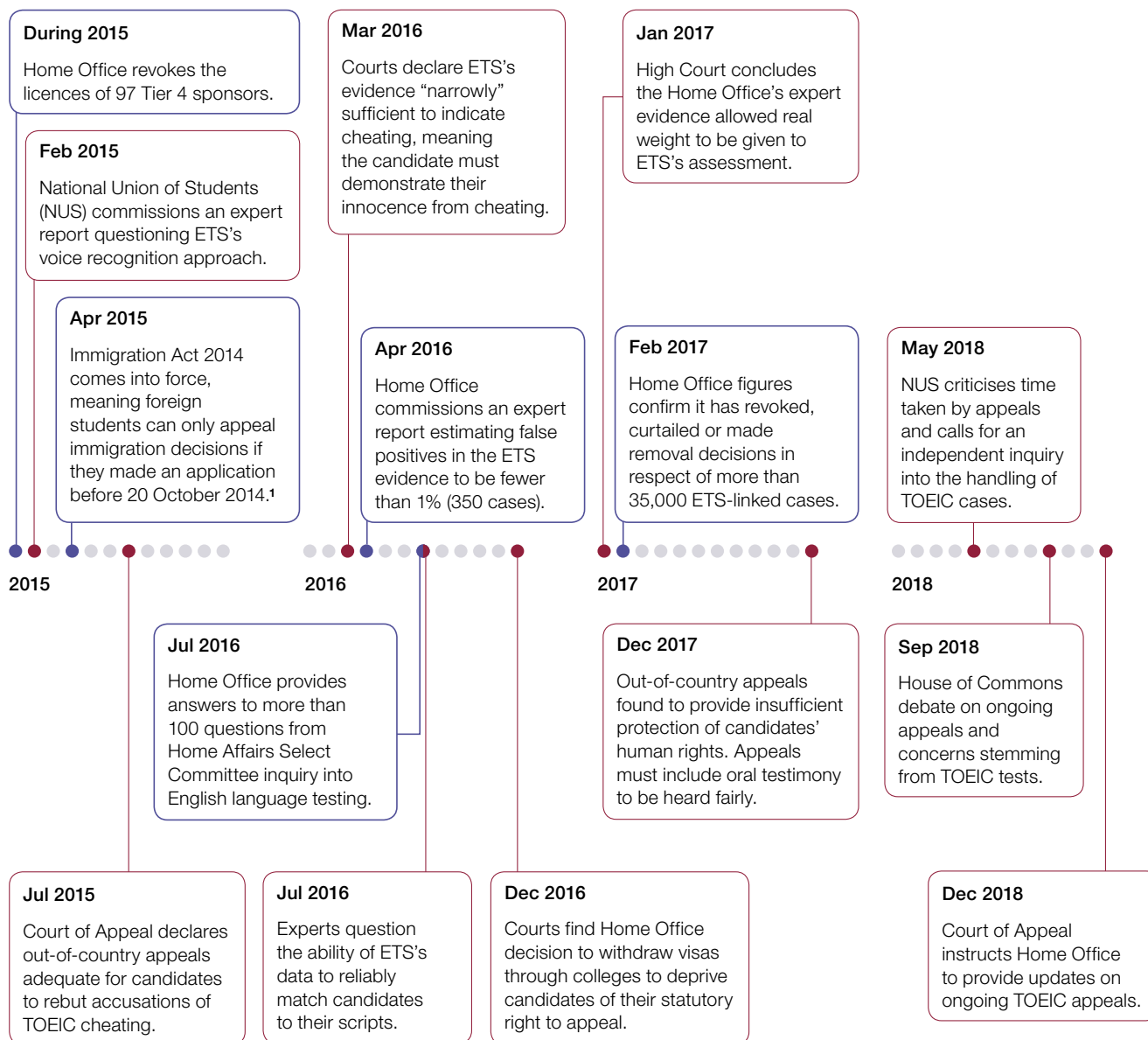
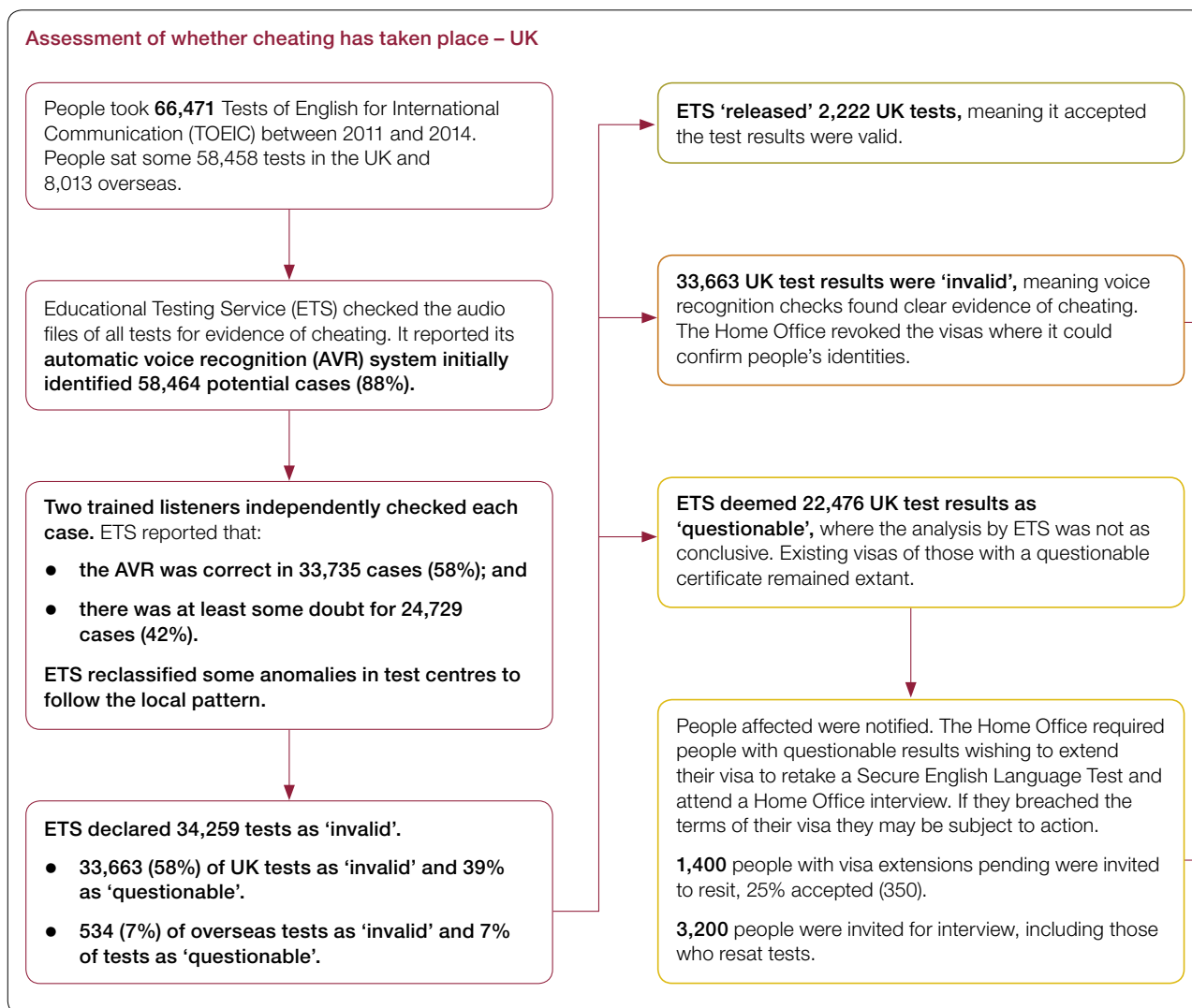


Figure 2

What happened to people taking Test of English for International Communication (TOEIC) exams

The Home Office took action against people Educational Testing Service (ETS) identified as cheating in tests



□ People with a questionable test certificate

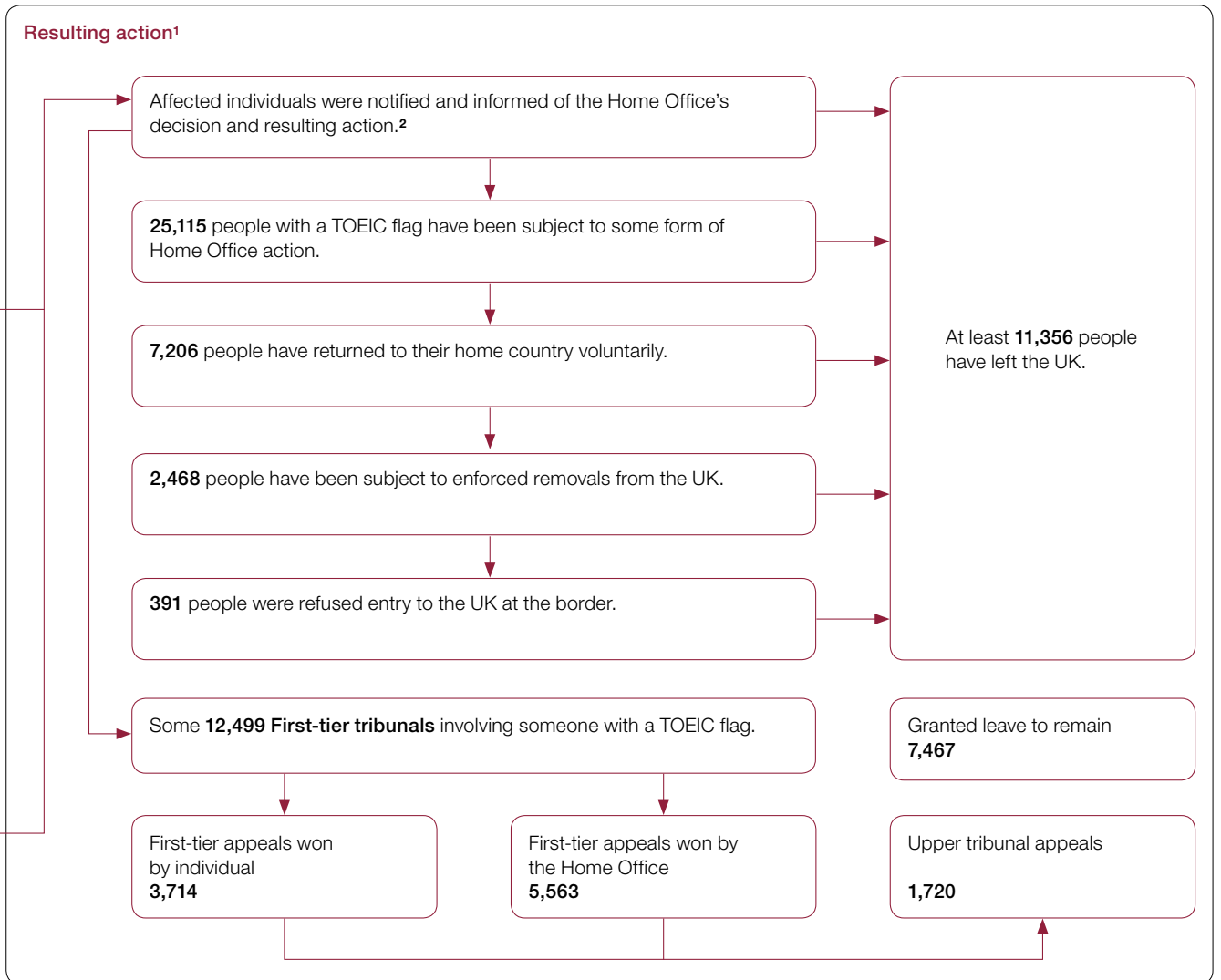
□ People with an invalid test certificate

□ People with a released test certificate

Notes

- 1 Numbers represent people with a TOEIC flag on their immigration record. Action may have been taken for reasons other than their TOEIC test result. Home Office action can include removing a person or refusing their application.
- 2 The number of people the Home Office took action against is as of March 2019 and should be treated as indicative due to data limitations.
- 3 ETS declared a test result as 'invalid' where it believed there was clear evidence of cheating. It declared a test result as 'questionable' where its analysis was not as conclusive and it was not confident in the test result (for example, because a test centre had a large number of invalid tests).

Source: National Audit Office analysis of public and Home Office data



Summary

Key findings

On the number of people identified as cheating

1 In 2014, the Home Office (the Department) obtained evidence of large-scale organised fraud from multiple sources. The Department already had longstanding concerns about the student visa system, and had taken action against hundreds of colleges. Following the *Panorama* programme in February 2014, the Department concluded that the system was too open to abuse and moved to investigate all Secure English Language Test (SELT) providers. It obtained data on exams and categories of cheating from US-based Educational Testing Service (ETS), the provider implicated by *Panorama*. It used this information to investigate sponsors, test centres and individual students. It found widespread failings in sponsors and test centres, including evidence of organised criminality. The Department had other data and intelligence which pointed to large-scale deception within the system (paragraphs 1.13 and 1.14, 2.1, 2.2 and 3.8).

2 ETS identified virtually every test in the UK as suspicious. It identified that 97% of all UK tests were invalid or questionable. ETS had not used voice recognition technology before to detect cheating in Test of English for International Communication (TOEIC) exams. As a precaution, ETS assigned two human listeners, based in the US, to each independently verify the results of each test. They confirmed that 58% of all tests were 'invalid' and marked the rest as either 'questionable' or 'cleared of suspicion'. ETS then reviewed the overall patterns for each test centre and reclassified some cases as invalid, questionable or cleared of suspicion. It identified 58% of 58,361 UK test participants as having invalid results and 39% as questionable. The Department could not tell us how many tests ETS reclassified following initial checks. ETS's lists suggest that it classified 462 more tests as invalid than the total number confirmed by human listeners (paragraphs 2.5 and 2.6).

3 For two years the Department revoked the visas of anyone with an invalid test, without expert assurance of the validity of voice recognition evidence.

The Department normally relies on SELT providers for assurance when awarding visas. It visited ETS, the US test provider, in June 2014 to understand the process and listen to some recordings. Based on legal advice at the time the Department considered ETS's evidence was sufficient to take action against students without the need for expert review. To support an appeal in 2015 the National Union of Students (NUS) commissioned an expert review, which said that voice recognition software could have made errors in up to 20% of cases and human listeners could have made errors in up to 30% of cases. On hearing the evidence, the courts criticised the Department staff's lack of expertise in challenging ETS's assessment but still concluded that the Department's overall evidence was sufficient to support its allegations. The Department instructed an eminent expert on speaker identification in 2016 to form an independent opinion on how likely it was that someone ETS identified as invalid may in fact be the innocent victim of a false positive (paragraphs 2.8 to 2.10).

4 In 2016, the Department's independent expert estimated that voice recognition checks would have identified substantially fewer than 1% of people of cheating incorrectly, based on a series of assumptions.

ETS faced a considerable workload to check the voice recordings from 66,000 tests. Initially 19 assessors experienced in voice biometrics had worked intensively carrying out checks. Voice recognition technology had not been used before with TOEIC tests and ETS did not use piloting or a control group. The expert judgement is based on more information than the expert the NUS commissioned, but it still needed to make a series of assumptions to estimate how many people may have been incorrectly identified. The estimate begins with the assumption that an error rate of 2% from a pilot with a similar ETS language test, the Test of English as a Foreign Language (TOEFL), also applied to the TOEIC tests. The expert said that because additional human checks rejected 42% of results it was likely that the number of innocent people would be very substantially less than 1%, with substantially more people given the benefit of the doubt. Neither expert had all the information they wanted to make their assessment. They did not know what software had been used, have access to many voice recordings, or know the performance of human verifiers. A judicial review in January 2017 accepted that the Department's expert opinion provided support for ETS's assessment. However, representatives of the Department conceded that neither the expert evidence nor ETS's assessment of invalidity could prove on their own whether a claimant cheated (paragraphs 2.11 to 2.14).

5 Most but not all people identified as cheating by ETS had very high marks.

Many people found to have invalid tests had scores indicating most were very fluent English language speakers. The distribution of scores indicates 92% (30,000) were fluent English speakers, of which 16,000 (49%) were highly fluent. People with questionable tests also had markedly skewed results. People cleared by voice recognition checks had a normal distribution of marks, similar to that of other English language tests (paragraph 2.16 and Figure 5 on pages 26 and 27).

6 It is difficult to estimate accurately how many people may have been wrongly identified. The degree of error is difficult to determine accurately without piloting or a control group for TOEIC tests. Analysis of the distribution of marks can help to identify anomalies compared with similar tests and people whose scores are not easily explained by known methods of cheating. We identified anomalies which the Department had not investigated. For example:

- In speaking tests deemed invalid, 100 people (0.3% of invalid results) had lower scores than the level required for study in the UK, meaning supposed proxies were actually people with limited English language ability.
- Thousands of people suspected of cheating by voice recognition checks had low scores in multiple choice tests. *Panorama* had shown evidence that test centres read out multiple choice answers to all candidates sitting a test. ETS cleared only 2,200 people outright of cheating but 5,900 multiple choice reading test results and 3,500 listening tests did not meet the level required for study, suggesting they were not given the answers.
- ETS cleared only 93 people with the very highest levels of fluency. In similar English language tests, a greater number of highly fluent speakers score top marks, but this pattern was not evident for people cleared of cheating (paragraphs 2.18 and 2.19).

7 Thousands of people accused of cheating have still won the right to remain in the UK. The courts have taken a case-by-case approach, having accepted that the Department's evidence places the burden of proof on individuals. Individuals' testimonies can be the deciding factor, and data indicate that thousands have succeeded in obtaining immigration status despite the Department's decision to revoke their visa for deception in TOEIC tests. People flagged with invalid test results on the Department's systems have:

- won 2,664 First-tier appeals against immigration decisions since 2014, equivalent to 8% of all those with invalid test results; people usually had to appeal on human rights grounds because they could not appeal the decision directly; and
- gained leave to remain in the UK in 4,157 cases: 532 have indefinite leave to remain, 477 are now British citizens and 2,682 have leave to remain of more than a year. Numbers represent people with some form of leave to remain as at March 2019 but may not represent the overall number of people that gained leave, because they are based on a reporting tool originally designed in 2014 and it may not capture some changes to people's immigration status (paragraphs 2.9, 3.30 and Figure 11).

8 We saw no evidence that the Department considered whether ETS had misclassified individuals or looked for anomalies. The Department provided us with evidence of the steps it had taken to analyse ETS data to check for errors and explained the processes it followed to correctly identify people on Home Office systems. It had not investigated the reasons why people with invalid scores had low marks, won appeals or gained leave to remain. The Department suggested that people only required low scores for some visas, such as for civil partnerships, and that it had used statistical analysis as far as is possible to check for errors. It said it would have needed to do a manual review of case files to reach a definitive view. The Department told us it could not provide accurate numbers on appeals which people won specifically against allegations of TOEIC deception because it had not carried out significant analysis on this since 2016, and HM Courts and Tribunals Service records these as human rights appeals. The Department said it investigated issues raised at appeal and that these were specific to each case. It does not see a significant risk of innocent people being wrongly caught up in action because the 2016 independent expert report and its other evidence indicated the number of errors was very low (paragraphs 2.17 and 3.32).

9 It was not possible for the Department to directly check the accuracy of ETS classifications. The Department has not had access to check ETS assessments of cheating. Instead it analysed the distribution of results by test centre and by test session and compared these with scores from other SELT providers. It concluded that ETS had not made systematic errors. We reviewed ETS's data and identified one systematic error. Some 6,000 people had been incorrectly categorised as 'UK nationals'. The Department did not know the reason for this and told us it might have been down to individuals stating their nationality incorrectly, test centres faking results, or IT systems defaulting to UK nationalities where no entry is recorded. The error does not demonstrate whether other data in the ETS lists are correct or not. In July 2016 experts for the Department and defendants said that they could not be completely confident in ETS because audio files did not have the original time, date and location stamps to enable cross-checking. In 2018 the courts considered ETS evidence on the security of data as part of an overall package of evidence that proved sufficient for criminal convictions (paragraphs 2.23 to 2.25).

On action taken against people accused of cheating

10 The Department estimated it had spent £21 million dealing with TOEIC as at November 2017. The Department did not keep detailed records of its spending but put together an estimate as part of preparations for possible legal action in 2017. It estimates it spent £9 million on immigration enforcement, £5.5 million on appeals and judicial reviews, £3 million dealing with temporary visa applications and £2 million on criminal and financial investigations. The Department made £624,000 of confiscations and cash seizures (paragraph 3.7 and Figure 8 on page 33).

11 The Department took a number of actions against education and testing institutions, including ETS. In 2014, the Department revoked the sponsor licences of 75 colleges and one university. It suspended all TOEIC test centres and allowed the ETS testing licence to expire in April 2014. In 2017, the Department concluded a three-year investigation to determine whether ETS and/or any of its employees were criminally complicit in undermining the TOEIC test. It did not find evidence of complicity. The Department also launched civil proceedings in 2017 against ETS Global BV (who administered the tests in the UK) rather than ETS US (who provided the analysis that the Department relied on to take action). The Department reached a financial settlement with ETS Global BV in 2018 for £1.6 million after legal advice that it did not have significant assets, and that Department estimates were extremely vulnerable in successfully demonstrating costs incurred. The Department told us decisions, both to initiate litigation and to reach settlement (in early 2018), were taken in consultation with ministers and in line with legal advice that it represented the best value for the taxpayer (paragraph 3.12 and 3.16).

12 The Department has supported successful prosecutions against 25 organised criminals. So far, 31 defendants from six organised crime groups have stood trial and 25 have been convicted. Twenty-one have received prison sentences totalling 70 years, and four more await sentencing. All those identified as having planned and organised the offences have been convicted. Fourteen further suspects are awaiting a decision on charging. The Department investigated 20 colleges in total, which it prioritised based on evidence from ETS and other information (paragraphs 3.8 to 3.13).

13 Students at the colleges that had their licences revoked were affected irrespective of whether they sat TOEIC or another English language test. Students had to find courses elsewhere if their college had its Home Office licence revoked. Some have struggled to find sponsors. The Department set up a sponsorship working group to look at how students could be supported to find other sponsors. It established an online tool containing details of 62 sponsors with courses available and wrote to 4,795 students in need of assistance. Some 837 students used the portal between November 2014 and January 2015. Stakeholders including Universities UK, the NUS and the UK Council for International Student Affairs (UKCISA) raised concerns about the effectiveness and progress of the sponsorship working group (paragraphs 3.20 to 3.22).

14 The Department needed to identify thousands of individuals on its systems before it could take action against them.

As at February 2016 the Department had identified 51,000 potential matches on its system. Further case work reduced the number of potential matches to around 42,000, with the Department confirming it had accurately matched names on its systems to ETS lists in 16,000 cases and a further 26,000 matches unconfirmed. The Department issued letters to all people with invalid tests saying their visa had expired and they would need to leave the country immediately. People with questionable results were told they would need to retake the test if they wished to extend their visa (paragraphs 3.24 and 3.25).

15 As at the end of March 2019 data showed that at least 11,000 people who had taken TOEIC tests had left the country following detection of mass fraud.

The number is based only on 11,000 invalid and 5,000 questionable cases with confirmed matches and people may have left for reasons not related to allegations of deception in TOEIC tests. Around 7,200 left voluntarily after April 2014. Around 2,500 people were forced to leave and almost 400 were refused re-entry to the UK on arrival at a port. UKCISA gave us evidence of several students detained at the border in relation to taking a SELT in the past or their English language ability. The Department told us it would only have taken action against people with an invalid certificate or other evidence against them (paragraph 3.28 and Figure 10 on page 37).

16 Around 12,500 appeals involving individuals matched to invalid or questionable TOEIC certificates have been heard up to March 2019.

The Immigration Act 2014 removed the right of appeal against applications for leave to remain in the UK. Students could challenge refusals of leave through judicial review. Where students made a separate human rights claim to the Department, and that claim was refused, the student could lodge a human rights appeal. Some people refused under the previous appeals system only had an out-of-country right of appeal. Stakeholders told us some people left the UK only to find they lacked access to the legal skills needed for UK courts. In December 2017, the courts ruled that out-of-country appeals did not offer sufficient protection for human rights and that appeals must be heard in the UK. The Department is investing in video-conferencing so that out-of-country appeals can be heard in future (paragraphs 3.30 and 3.31, 3.36 and 3.37).

17 Individuals matched to invalid or questionable TOEIC certificates won 3,600, First-tier appeals but the Department has not tracked the reasons why. In most cases, appeals were made on human rights grounds. Individuals suspected of cheating won some 40% of First-tier appeals heard in court compared with 60% won by the Department. The Department only analysed the reasons why it lost the first 385 appeals up to September 2016, because case law was not yet established. In 17% of these cases people winning were proven innocent of cheating in TOEIC tests and in 57% cheating was not proven. The Department does not hold readily available data on how many people proved their innocence of cheating in TOEIC cases and how many people won for other reasons. A detailed review of the case files would be required to establish this. Human rights grounds, such as the right to family life, will have become more significant with the passage of time. The most recent available data for 28 appeals in May 2018 indicated almost all appellants won where they had a British child or spouse. Appellants without British relatives tended to be unsuccessful (paragraphs 3.32, 3.34 and 3.35 and Figure 11 on page 39).

Concluding remarks

18 Evidence from ETS, Departmental analysis and successful prosecutions against those involved in administering test centres strongly suggests that there was widespread abuse of the Tier 4 visa system. The exact scale of that abuse is difficult to determine because experts drew different conclusions over the reliability of voice recognition checks. It may be right to say that the Department expert was more informed than the NUS expert, but not that he was sufficiently well informed to provide a definitive view. The Department nonetheless relied on his judgement to assume that collateral damage would be low. The data environment is weak and the Department has not been able to independently validate ETS judgements about cheating. It is reasonable based on the balance of probabilities to conclude there was cheating on a large scale because of the unusual distribution of marks, and high numbers of invalid tests in test centres successfully prosecuted for cheating.

19 The Department decided to react vigorously to the evidence of widespread cheating that was reported by *Panorama*, and which confirmed their pre-existing concerns. Widespread action to close colleges meant students who did not sit TOEIC exams have been affected, some of whom have struggled to find replacement sponsors. The courts found in 2016 that the Department's evidence was sufficient to make the accusations it did. Nonetheless, the Department's course of action against TOEIC students carried with it the possibility that a proportion of those affected might have been branded as cheats, lost their course fees, and been removed from the UK without being guilty of cheating or adequate opportunity to clear their names. To date the Home Office has considered its evidence sufficient to tell people that the number of people wrongly affected would be very small, and that they can appeal from outside the UK or lodge a Judicial Review.