



National College for  
Teaching & Leadership

# **Linking ITT and workforce data: (Initial Teacher Training Performance Profiles and School Workforce Census)**

**Research report**

**July 2016**

**Rebecca Allen, Dave Bibby, Meenakshi  
Parameshwaran and Philip Nye – Education  
Datalab, FFT**

# Contents

List of figures	3
List of tables	3
Purpose of the report	4
Key findings	4
Estimating retention rates by route	6
Lower bound retention rates by route	7
Variation in retention rates by route and individual characteristics	10
Upper bound retention rates	13
Technical Annex: Matching SWC and ITTPP	16
Consistently identifying teachers across datasets	16
Using the Teacher Reference Number as an identifier across datasets	16
Data availability for fuzzy matching of teachers	17
Fuzzy matching procedure and performance	17
Missing records in SWC	20
Known missing records for employment-based ITT teachers	20
Finding missing records using the date of arrival in school	21

## List of figures

Figure 1 Upper and lower bound retention rates by route.....	15
--	----

## List of tables

Table 1 Lower bound estimates of retention by route	9
Table 2 Retention rates by route for women aged 24 and under	11
Table 3 Odds ratio of remaining in teaching 3 months and 2 years 3 months after QTS	13
Table 4 The FFT ID in SWC	18
Table 5: Patterns of presence across years of SWC	19
Table 6 Identifying missing records using date of arrival in school	22
Table 7: Success rate in identification of Teach First records in School Workforce Census	22

## Purpose of the report

This report gives some initial estimates of retention in the state-funded teaching workforce in England by teacher training route, as a proportion of all those first registering on an ITT course. We illustrate how this varies by region and teacher characteristics. We give lower and upper bound retention rate estimates, reflecting uncertainty inherent in the data.

The technical annex discusses data quality issues in the Initial Teacher Training Performance Profiles (ITTPP) and the School Workforce Census (SWC) that are particularly relevant to estimating early career teacher retention. (These issues do not detract from the use of the survey to derive statistics for the workforce as a whole.)

## Key findings

Given imprecision in the estimates in this report it is important not to over-interpret relatively small differences in retention. However, a few findings are striking enough that they are unlikely to diminish as data quality improves. These are:

- Three regions of England – North East, North West and South West – appear to have large numbers of new qualified teachers who do not join a state-sector school immediately after achieving QTS.
- Those studying on undergraduate with QTS courses have low initial retention rates in the profession, though we cannot know whether this results from subsequent choices made by the individual or recruitment decisions made by schools.
- Teach First has very high two year retention rates, but thereafter their retention is poorer than other graduate routes.
- Ethnic minority teacher trainees have very low retention rates.
- Individuals who train part-time or who are older have much poorer retention rates, which may simply reflect other family commitments that interfere with continuous employment records.

Note that two parallel independent reports are being published which present more detailed analysis of teacher retention drawing on the same data. The production of these additional reports has been made possible because the data was released to the report authors under the terms of a contract with the Department for Education (DfE), but the analysis and conclusions included in the additional reports have been carried out independently of DfE to meet the requirements of Teach First and the Nuffield Foundation. DfE has not been involved in the analysis underpinning, or the writing of, these two additional reports.

The report commissioned by Teach First is entitled: 'The careers of Teach First Ambassadors who remain in teaching', by Rebecca Allen, Meenakshi Parameshwaran and Philip Nye. The report commissioned by the Nuffield Foundation is entitled 'The Longer-Term Costs and Benefits of Different Initial Teacher Training Routes' by Rebecca Allen, Chris Belfield, Ellen Greaves, Caroline Sharp and Matt Walker. The lead researcher for this project is Ellen Greaves at the Institute for Fiscal Studies.

## Estimating retention rates by route

Estimates of early career retention rates by route are reported here on the basis of those found each year in state maintained schools in England, as captured by the School Workforce Census (SWC), as a percentage of the total number starting the training route according to the Initial Teacher Training Performance Profiles (ITTPP). This reflects the social investment in training teachers; it is important that we capture drop-out prior to the achievement of QTS. This does mean that all the statistics in the report will differ substantially from those published elsewhere by the Department for Education, which typically report retention in the workforce as a proportion of those awarded QTS.

It is not possible for us to straightforwardly decompose our statistics into drop-out prior to and post-QTS since we use a slightly different sample of teachers reflecting the fact that a small proportion of trainees change route or course en route to being awarded QTS. In our method, if a trainee begins an HEI-led full-time PGCE in 2010 and drops-out and then registers on an employment-based route starting 2011, achieving QTS in 2012, we will classify this individual as following HEI-led PGCE with an expected QTS date of 2011. We will indicate that they achieve QTS, even though it is via a different route and will attribute any appearances in the SWC to them (though they will not appear in Year 1 since this would be their HEI-led training year where they are not an employee in school). By contrast, most statistics will count this as two individuals: both are included in statistics indicating the percentage achieving QTS; the second is included in post-QTS retention rates from the date of actual QTS in 2012.

The School Workforce Census is an annual census of all those working in state maintained schools each year from November 2010 onwards. We use the first five sweeps of the Census in our analysis here. The data is compiled by local authorities, academy chains or individual schools from their Management Information Systems and includes basic demographic, role and pay information on staff. SWC is used by the Department for Education to produce workforce statistics that support teacher supply planning.

The Initial Teacher Training Performance Profiles are used by the National College for Teaching and Leadership to collect information on teacher trainees from initial teacher training providers. The data includes information about the number of trainees, basic demographic information and their qualification on entry to their ITT course.

We categorise each ITTPP record into the following routes:

- HE provider-led, i.e. the most common post-graduate training route leading to PGCE as well as QTS. We split trainees into those studying full-time and part-time;
- SCITT, the School Centred Initial Teacher Training programme, where a school is registered as the training partner, which may lead to award of PGCE as well as QTS;

- EBITT, which incorporates all Employment-Based Initial Teacher Training programmes prior to the introduction of School Direct (principally the Graduate Teacher Programme);
- School Direct Salaried is a new employment-based training route, delivered in partnership with an initial teacher training institution;
- School Direct Training is a new employment-based training route, delivered in partnership with an initial teacher training institution, where the trainee is not an employee during their training year;
- Teach First is a two-year employment programme where QTS is awarded at the end of the first year;
- Undergraduate with QTS incorporates all HEI undergraduate programmes leading to QTS;
- Assessment Only award of QTS is assigned to those following a programme leading to QTS who have previous experience of teaching (e.g. overseas teachers). We do not include these in the analysis tables since they are small in number.

We provide a further re-categorisation of records using the programme title field to search for keywords such as “School Direct”, “TF”, “Salaried”, etc... For each record we calculate the year that QTS should normally have been achieved. We heavily rely on the commencement date and length of course to do this, but the data has a large number of inconsistencies between fields and so we also re-classify where course lengths appear unreasonable. Our expected QTS date field is less likely to be accurate for part-time trainees and for those following an undergraduate with QTS course where length of course varies across institutions.

Where individuals have multiple records we favour the earliest expected completion date, and the earliest ITTPP record date where required. If an individual still has multiple records then we favour that relating to the employment based route.

## Lower bound retention rates by route

We treat the SWC estimates of retention rates as lower bound estimates, knowing that we may be failing to correctly match some teachers across datasets and knowing there are missing records in SWC. Table 1 reports these lower bound estimates by route and expected QTS year, for full-time trainees only apart from the part-time HEI-led route that we report separately (there are too few part-time trainees in other routes to disaggregate). The data on the pre-QTS year relates to the presence of records in SWC during the (final) training year. It is worth noting that we find HEI-led students, BA with QTS students and SCITT students in SWC where they should not be during the year. Occasionally this may be because we are wrongly writing back records based on arrival in school dates. In the case of the part-time PGCE students, they may be working as

unqualified teachers whilst completing their PGCE or they may have completed it quicker than the course length field indicates they should. We find those on employment-based routes – EBITT, School Direct salaried and Teach First – at much lower rates than we should in their pre-QTS year where they should all be employed at a school (there is very little course drop-out by census date in November).

Since these estimates are subject to significant uncertainty arising from data quality it is important not to over-interrogate them. But roughly speaking the following patterns emerge:

- School Direct Training, including the salaried route, EBITT and SCITT all look quite similar in having relatively high retention rates
- The full-time HEI-led and the undergraduate with QTS routes have similar retention rates by year 3 but the initial profile looks quite different. The undergraduate with QTS has low proportions teaching in year 1 suggesting either:
  - New teachers may first be taking a gap year before returning to teach in a state school in England
  - We have misclassified the course length or they are taking an additional year to complete and QTS is not yet achieved
  - They are disproportionately affected by missing records in their first year of teaching for some reason
- The part-time HEI-led PGCE has relatively poor retention rates. We do not look at retention for part-time students following other routes but it is likely to be similarly low. Clearly those choosing to, or needing to, follow a programme part-time will often have other life commitments that will interfere with a continuous employment record.
- Teach First has high retention for the period of their initial programme (pre-QTS and year 1 in the table) and rapidly falling retention thereafter.



Table 1 Lower bound estimates of retention by route

		N	Pre-QTS year	Year 1	Year 2	Year 3
<b>HEI-led (FT)</b>	<b>2010</b>	21,477		58%	61%	60%
	<b>2011</b>	21,242	1%	55%	62%	61%
	<b>2012</b>	19,681	1%	62%	66%	61%
	<b>2013</b>	19,427	1%	65%	66%	
	<b>2014</b>	17,413	1%	67%		
<b>HEI-led (PT)</b>	<b>2010</b>	1,250		41%	43%	44%
	<b>2011</b>	1,570	21%	41%	45%	44%
	<b>2012</b>	1,568	21%	50%	52%	49%
	<b>2013</b>	1,540	18%	50%	51%	
	<b>2014</b>	1,080	21%	48%		
<b>Undergraduate with QTS</b>	<b>2010</b>	6,859		51%	61%	64%
	<b>2011</b>	7,380	4%	42%	55%	58%
	<b>2012</b>	7,569	3%	48%	59%	58%
	<b>2013</b>	7,420	3%	51%	57%	
	<b>2014</b>	6,950	3%	54%		
<b>EBITT</b>	<b>2010</b>	6,314		75%	71%	69%
	<b>2011</b>	5,748	84%	73%	72%	69%
	<b>2012</b>	6,062	80%	75%	73%	69%
	<b>2013</b>	4,997	79%	75%	70%	
	<b>2014</b>	236	80%	73%		
<b>SCITT</b>	<b>2010</b>	2,114		64%	66%	65%
	<b>2011</b>	2,024	5%	60%	68%	67%
	<b>2012</b>	2,108	7%	68%	70%	67%
	<b>2013</b>	2,355	6%	69%	69%	
	<b>2014</b>	2,388	3%	67%		
<b>School Direct Salaried</b>	<b>2010</b>	0				
	<b>2011</b>	0				
	<b>2012</b>	0				
	<b>2013</b>	0				
	<b>2014</b>	2,375	70%	66%		
<b>School Direct Training</b>	<b>2010</b>	0				
	<b>2011</b>	0				
	<b>2012</b>	0				
	<b>2013</b>	371	4%	70%	67%	
	<b>2014</b>	4,049	8%	70%		
<b>Teach First</b>	<b>2010</b>	502		82%	55%	46%
	<b>2011</b>	568	89%	84%	57%	48%
	<b>2012</b>	765	89%	80%	57%	43%
	<b>2013</b>	1,001	90%	83%	55%	
	<b>2014</b>	1,274	81%	73%		

## Variation in retention rates by route and individual characteristics

Teacher retention varies by route, in part because the different training experiences alter the propensity of teachers to decide to stay in the profession and in part because different routes attract and accept different types of individuals. It is important to recognise this distinction in policy terms, but it isn't possible to disentangle the two in non-experimental data. However, we can report how propensity to stay in the profession varies by route, holding constant a set of individual characteristics. There are a number of ways we can illustrate this. Table 2 reports retention rates for women teachers aged 24 and under, the most prevalent group of new teachers. Table 3 shows the results of two logistic regressions that report how retention rates vary by route, holding constant individual characteristics.

There are a number of patterns that emerge when we explore retention rates for the women aged 24 and under in Table 2:

- Retention rates are higher for this group than for the population of new teachers as a whole
- This reveals large differences in staying on rates between young females completing the HEI-led full-time and those completing the undergraduate with QTS.

Table 2 Retention rates by route for women aged 24 and under

		N	Pre-QTS year	Year 1	Year 2	Year 3
<b>HEI-led (FT)</b>	<b>2010</b>	12,614		64%	67%	66%
	<b>2011</b>	12,484	1%	60%	68%	67%
	<b>2012</b>	11,927	1%	68%	72%	67%
	<b>2013</b>	12,001	1%	71%	71%	
	<b>2014</b>	10,911	1%	73%		
<b>HEI-led (PT)</b>	<b>2010</b>	507		44%	48%	49%
	<b>2011</b>	574	20%	46%	50%	48%
	<b>2012</b>	556	20%	56%	57%	56%
	<b>2013</b>	576	17%	60%	62%	
	<b>2014</b>	377	18%	54%		
<b>Undergraduate with QTS</b>	<b>2010</b>	5,423		52%	62%	65%
	<b>2011</b>	5,627	3%	45%	59%	63%
	<b>2012</b>	5,951	2%	50%	60%	60%
	<b>2013</b>	5,796	2%	54%	60%	
	<b>2014</b>	5,456	2%	57%		
<b>EBITT</b>	<b>2010</b>	2,827		77%	74%	71%
	<b>2011</b>	2,630	84%	75%	73%	70%
	<b>2012</b>	2,670	81%	78%	76%	71%
	<b>2013</b>	2,361	80%	79%	74%	
	<b>2014</b>	125	81%	73%		
<b>SCITT</b>	<b>2010</b>	1,126		69%	73%	72%
	<b>2011</b>	1,106	4%	65%	72%	72%
	<b>2012</b>	1,201	7%	75%	75%	72%
	<b>2013</b>	1,352	5%	75%	75%	
	<b>2014</b>	1,376	3%	73%		
<b>School Direct Salaried</b>	<b>2010</b>	0				
	<b>2011</b>	0				
	<b>2012</b>	0				
	<b>2013</b>	0				
	<b>2014</b>	951	72%	70%		
<b>School Direct Training</b>	<b>2010</b>	0				
	<b>2011</b>	0				
	<b>2012</b>	0				
	<b>2013</b>	192	5%	76%	71%	
	<b>2014</b>	2,267	7%	75%		
<b>Teach First</b>	<b>2010</b>	324		80%	56%	49%
	<b>2011</b>	331	91%	88%	62%	53%
	<b>2012</b>	461	89%	82%	58%	46%
	<b>2013</b>	631	90%	86%	59%	
	<b>2014</b>	836	82%	74%		

Table 3 shows the output from two logistic regressions estimating the odds ratio of remaining in teaching 3 months after the training course ends (i.e. when QTS should be achieved) and 2 years 3 months after QTS is achieved. Once again, it does so for all trainees starting a teacher training course for the first time, regardless of whether they drop out during the course or not. It allows us to explore retention rates, holding constant the individual and regional characteristics of the trainees, and also to explore how these factors affect retention overall.

Note that overall the explanatory power of these regression models is very low, which means that route, individual and regional factors are relatively poor predictors of whether an individual leaves the profession. However, the following patterns do emerge:

- Teach First and EBITT have the highest retention rate 3 months after QTS. This period forms part of Teach First's initial placement; note that Teach First retention is lower than any other route by 2 years 3 months after QTS. SCITT and the new employment-based routes have slightly higher retention than the full-time HEI-led route. The undergraduate with QTS and part-time HEI-led have low retention rates.
- Initial retention rates appear to have risen a little over the period 2010 to 2014, though we need to be a little careful in this assertion since it may be affected by differences in data quality between SWC or ITTPP sweeps. Retention rates at 2 years 3 months appear to be more stable. Rising retention rates are plausible despite a recovering economy if demand for teachers is rising or because numbers trained have fallen.
- Early career retention rates are significantly higher for women and are significantly lower for ethnic minority teachers.
- Year 1 retention rates are lower in all regions of training outside London, except for the East of England. Retention rates are particularly low in the North East, the North West and the South West. These patterns are consistent with known demand for teachers.
- Retention rates fall consistently by age of new teacher.
- Initial retention rates are higher for English, mathematics and primary teachers than they are for science and other secondary subject teachers, but by the start of the third year of teaching the retention rate of mathematics secondary teachers has fallen to match levels seen in science and other secondary subjects.

**Table 3 Odds ratio of remaining in teaching 3 months and 2 years 3 months after QTS**

	3 months after QTS			2 years 3 months after QTS		
	OR	S.E.		OR	S.E.	
<b>Vs HEI-based FT:</b>						
HEI-based PT	0.70	(0.02)	***	0.68	(0.02)	***
U'grad with QTS	0.51	(0.01)	***	0.68	(0.01)	***
EBITT	2.03	(0.04)	***	1.55	(0.03)	***
SCITT	1.24	(0.03)	***	1.23	(0.04)	***
School Direct salaried	1.08	(0.05)	*			
School Direct Training	1.19	(0.04)	***			
Teach First	1.94	(0.08)	***	0.46	(0.02)	***
<b>Vs expected QTS in 2010:</b>						
QTS in 2011	0.83	(0.01)	***	0.99	(0.02)	n.s.
QTS in 2012	1.07	(0.02)	***	0.95	(0.01)	***
QTS in 2013	1.18	(0.02)	***			
QTS in 2014	1.23	(0.02)	***			
<b>Female</b>	1.35	(0.02)	***	1.32	(0.02)	***
<b>BME ethnicity</b>	0.55	(0.01)	***	0.69	(0.01)	***
<b>Vs London region:</b>						
East Midlands	0.94	(0.02)	***	1.22	(0.04)	***
East of England	1.13	(0.03)	***	1.19	(0.03)	***
North East	0.60	(0.02)	***	0.89	(0.03)	***
North West	0.55	(0.01)	***	0.79	(0.02)	***
South East	0.84	(0.02)	***	0.92	(0.02)	***
South West	0.59	(0.01)	***	0.84	(0.02)	***
West Midlands	0.83	(0.02)	***	1.13	(0.03)	***
Yorkshire and Humber	0.83	(0.02)	***	1.11	(0.03)	***
<b>Vs age band 22 and under:</b>						
Age band 23-24	0.99	(0.02)	n.s.	0.92	(0.02)	***
Age band 25-29	0.86	(0.02)	***	0.73	(0.02)	***
Age band 30-34	0.66	(0.02)	***	0.57	(0.02)	***
Age band 35-39	0.54	(0.02)	***	0.51	(0.02)	***
Age band 40-65	0.41	(0.01)	***	0.39	(0.01)	***
<b>Vs subject other secondary:</b>						
English	1.82	(0.05)	***	1.45	(0.04)	***
Mathematics	1.36	(0.03)	***	1.05	(0.03)	*
Science	0.93	(0.02)	***	0.93	(0.02)	***
Primary	1.11	(0.02)	***	1.32	(0.02)	***
<b>Constant</b>	<b>1.97</b>	<b>(0.05)</b>	<b>***</b>	<b>1.82</b>	<b>(0.06)</b>	<b>***</b>

Number of observations = 181,269; 110,320; Pseudo R-squared 6%; 3%.

## Upper bound retention rates

We know data quality problems lead us to understate teacher retention estimates in SWC. If we assume that internal records held by Teach First themselves are largely

accurate, they imply that we are understating retention by around 6 percentage points for participants on their programme. It may or may not be reasonable for us to extrapolate this estimate across other training routes, thus we have a great deal of uncertainty about true retention rates.

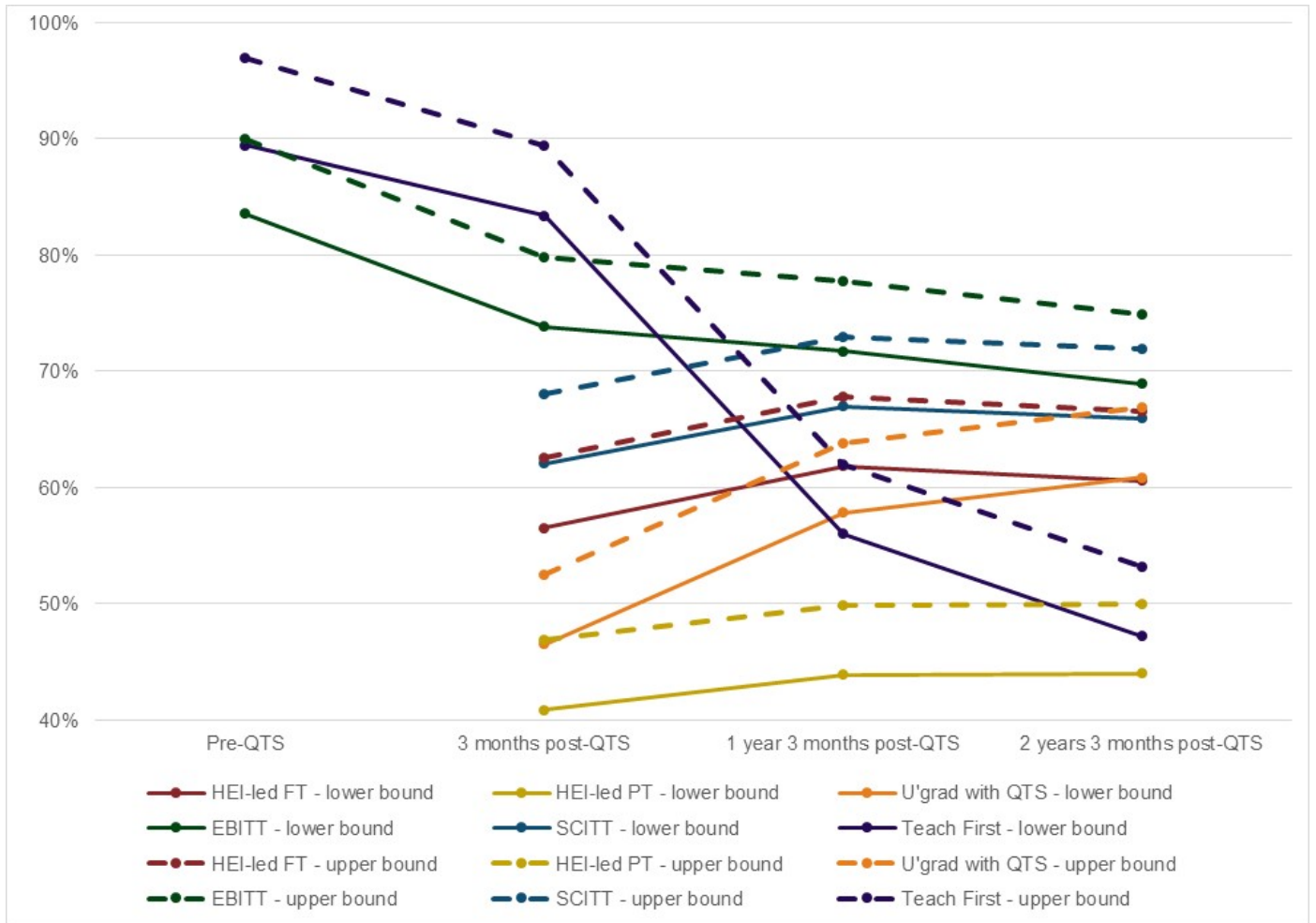
Our understanding of the nature of missing records in SWC leads us to believe the difference between the upper and lower bound of our teacher retention estimates should shrink as careers progress, but in Figure 1 we take the conservative approach to retaining it at 6% throughout the first few years of a teacher's career.

The chart illustrates a lower bound (i.e. found in SWC) and upper bound (i.e. 6 percentage points higher) estimate of teacher retention by training route for cohorts working towards achieving QTS in 2011 and 2012. At first glance, it may appear to be a large number of overlapping lines. This should serve to emphasise that, given the uncertainty in our estimates of retention, it is hard for us to assert that there are substantial differences in retention by route. However, it is possible to say:

- 3 months post-QTS: Teach First has the highest retention, followed by EBITT, followed by HEI-led and SCITT, with undergraduate with QTS and HEI-led part-time having lowest proportions in teaching.
- 1 year 3 months post-QTS: the pattern is very similar to the previous year, except that Teach First retention is now as low as undergraduate with QTS.
- 2 years 3 months post-QTS: the pattern is again similar, but Teach First's retention is now at the same rate as the part-time HEI-led route. By contrast, those following an undergraduate with QTS route are now teaching at similar rates to those completing the full-time HEI-led PGCE.

The former register of teachers held by the General Teaching Council for England (GTCE) could be used as an alternative to the Teach First records to calculate an upper bound for the retention rates presented here. The GTCE database is likely to contain the vast majority of teachers, as registration was a legal requirement for all qualified teachers in maintained schools, pupil referral units and non-maintained special schools prior to 2012. Comparison of the GTCE and SWC for teachers that achieved QTS in 2010 implies an upper bound for retention 3.1 percentage points higher than the estimates from SWC. This is the approach used in the Nuffield report 'The Longer-Term Costs and Benefits of Different Initial Teacher Training Routes' (Allen, Belfield, Greaves, Sharp and Walker) for all routes other than Teach First (where the upper bound used above is taken), but this report chooses to present the most conservative estimate of the upper bound.

Figure 1 Upper and lower bound retention rates by route



# Technical Annex: Matching SWC and ITTPP

## Consistently identifying teachers across datasets

To estimate the retention in the profession of those training to be teachers, we need to:

1. Consistently identify individuals where they appear on multiple occasions in the ITTPP. For our analysis in this report we wish to record their first registration on a course leading to QTS and calculate the earliest date at which QTS was achievable. Individuals will have multiple records if:
  - They registered on a course lasting more than one year;
  - They registered on a one-year course, but took longer than a year to complete it;
  - They registered on multiple courses. In this case we take the route and timing details from the first registration but look across all courses to see whether QTS was ever achieved;
  - If a provider recorded them in a year by error (for example, not removing them after completion or withdrawal);
  - If two different providers registered the same individual (for example, a school-centred initial teacher training provider and a higher education institution who are both involved in their training).
2. Consistently identify individuals within the five years of SWC data available.
3. Successfully identify teachers from the ITTPP who go on to teach in state-maintained schools in England, finding them in the SWC.

## Using the Teacher Reference Number as an identifier across datasets

The Teacher Reference Number is a 7-digit identifier allocated by the Qualified Teacher Status and Induction (QTS-I) team within the Department for Education (DfE). It is used as a unique teacher identifier for many purposes; for example, a TRN is allocated to all trainees on state-funded Initial Teacher Training programmes, is used for identification for pensions purposes, and is assigned to overseas teachers who are awarded QTS by virtue either of their existing teaching qualifications (EU and Schengen countries plus Canada, USA, Australia and New Zealand) or following assessment for QTS in the UK.

It should, in theory, be a unique identifier for teachers, but is occasionally not (e.g. a teacher who is unqualified and opts out of the Teachers' Pension Scheme will not have a TRN; a teacher may be allocated two different TRNs for rare administrative reasons; the TRN is recorded with error by a school).

The TRN is present on almost all records in the ITTPP (missing on just 323 records or 0.09%). TRNs appear just once in 74% of cases, twice in 13% of cases, three times in



10% of cases and four or more times in 3% of cases. (These multiple entries are mostly correct repeat entries of the same teacher in multiple years of ITTPP.)

In the version of the SWC that we use here, TRNs are missing at the rate of about 1% in years 2010 to 2013 and at a rate of 2% in 2014. The missing TRNs in the SWC are more prevalent among the youngest teachers. This may be because they are unqualified teachers who are choosing not to contribute to the Teachers' Pension Scheme and so do not possess a TRN. However, we think this is relatively rare. More likely, we believe it is because they are new to the school and the school has not entered their TRN into their management information system by the time of the November census.

## **Data availability for fuzzy matching of teachers**

Fuzzy matching is a technique used to link records within and between databases where a unique identifier cannot be used to match perfectly. For that reason, matches may be less than 100 per cent perfect. We use it here to consistently identify individuals within each of our databases and between our databases.

The TRN remains a key identifier, but we now assume it may be coded with error or that individuals may hold several TRNs simultaneously. In addition we draw on personal identifiers across the databases. In the ITTPP the fields available to us are date of birth, first name, surname, gender and ethnicity. It does not include an alternative surname field. Where training route precedes employment in schools this is not a problem for linking to SWC post-QTS award; it may lead to under-matching if an individual changes their surname between working as an unqualified teacher and taking a training course, or if they take multiple courses and change their surname during the process.

SWC has the same personal identifiers as ITTPP, but we additionally have a former family name field. Unfortunately, the data is incomplete in other respects. First, the name fields are missing for about half the records in the 2010 census, leading to significant risk of under-matching for individuals joining the profession in this year. Second, personal identifiers are missing where the SWC has been patched with additional records from the Database of Teacher Records, which is compiled from pension records. This has happened where a school has entirely failed to make a census return (23,164 records in total across 2010 to 2013; 2014 SWC has not yet been imputed).

## **Fuzzy matching procedure and performance**

We draw records from all our databases and implement a fuzzy match to create an alternative consistent teacher identifier that we call FFT ID. At the start of this process we clean the name text fields to remove errant characters, prefixes, suffixes and so on to improve the chances of consistency across databases. We implement a typical fuzzy matching procedure that prioritises exact matches on fields such as TRN and DOB, and attempts matches on common variants of names. We attempt to avoid over-matching by ensuring that our procedure does not result in teachers working in multiple schools in one

year. At the end of this procedure we perform a standard set of under-matching and over-matching checks and manually adjust a small number of the FFT IDs accordingly.

We evaluate the performance of the FFT IDs created by our fuzzy matching procedure by comparing them to the TRNs in the databases. Overall they are in considerable agreement. According to the FFT ID there are slightly more trainees in the dataset overall but there are also notably more individuals with multiple records (245,842 FFT IDs in the ITTPP versus 243,838 TRNs). Conflicts occur in only 5,880 FFT ID individuals (i.e. 2.4% of all the individuals identified using the fuzzy matching).

The creation of the FFT IDs identifies a slightly different set of teachers achieving QTS in the ITTPP compared to the records identified by DfE. The DfE sub-set of NQTs contains 193,434 records for teachers achieving QTS; we think this includes a small number of duplicates and can identify only 193,389 unique teachers in this subset who were awarded QTS. However, we find 199,419 teachers with unique FFT IDs who achieve QTS in the full ITTPP database.<sup>1</sup>

The linked 2010-2014 SWC database that we are using in our analysis has been set-up to contain one record per individual per year, with the main contract being selected where teachers work at multiple schools. Our FFT ID identifies small additional numbers of occasions where a teacher is present in multiple schools in a year, but with different (or missing) TRNs (see Table 4). On the whole this indicates that the quality and consistency of the DfE matching work is good.

**Table 4 The FFT ID in SWC**

	2010	2011	2012	2013	2014
Total number of observations	485,445	483,023	488,768	492,631	497,484
Number with missing FFT IDs	0	0	0	0	0
Number of distinct FFT IDs within year	485,418	482,985	488,731	492,601	497,430
FFT ID appears once in year	485,391	482,947	488,694	492,571	497,376
FFT ID appears twice in year	54	76	74	60	108

Table 5 shows patterns of presence across the five years of SWC for teachers using the FFT ID and the TRN as alternative identifiers. The FFT ID has fewer occasions where an individual is identified as present for only one of the years of the census. In the majority of cases where the TRN identifies an individual in one year only and the FFT ID does not, this is because the TRN is missing or invalid (i.e. 9999999 or 0). In these cases the FFT ID identifies the individual in question anywhere between two and five times.

---

<sup>1</sup> It is important to note that the ITTPP is not the main record of whether QTS is awarded. DfE have an independent record in the form of the Database of Qualified Teachers (DQT) for this purpose.

The FFT ID also identifies a larger number of individuals who are present across all five years of the census (48 per cent of all teachers). It is worth noting that both the TRN as an ID and the FFT ID identify relatively large numbers who are missing in 2010 and present in 2011-2014, or vice versa. If the missing name fields for the 2010 census could be recovered then it is possible that some of these individuals could be correctly matched across all years.

**Table 5: Patterns of presence across years of SWC**

Years present in:					FFT ID		TRN ID	
2010	2011	2012	2013	2014	N	%	N	%
No	No	No	No	Yes	46116	7%	48529	7%
No	No	No	Yes	No	9523	1%	12156	2%
No	No	No	Yes	Yes	34855	5%	34180	5%
No	No	Yes	No	No	7272	1%	9438	1%
No	No	Yes	No	Yes	2320	0%	2304	0%
No	No	Yes	Yes	No	5208	1%	5104	1%
No	No	Yes	Yes	Yes	31278	5%	30983	5%
No	Yes	No	No	No	7493	1%	9415	1%
No	Yes	No	No	Yes	928	0%	922	0%
No	Yes	No	Yes	No	622	0%	606	0%
No	Yes	No	Yes	Yes	2280	0%	2266	0%
No	Yes	Yes	No	No	4515	1%	4299	1%
No	Yes	Yes	No	Yes	1296	0%	1267	0%
No	Yes	Yes	Yes	No	4294	1%	4155	1%
No	Yes	Yes	Yes	Yes	31374	5%	30884	5%
Yes	No	No	No	No	39218	6%	40153	6%
Yes	No	No	No	Yes	2106	0%	2118	0%
Yes	No	No	Yes	No	945	0%	943	0%
Yes	No	No	Yes	Yes	2661	0%	2663	0%
Yes	No	Yes	No	No	1660	0%	1677	0%
Yes	No	Yes	No	Yes	517	0%	519	0%
Yes	No	Yes	Yes	No	1092	0%	1093	0%
Yes	No	Yes	Yes	Yes	7036	1%	7070	1%
Yes	Yes	No	No	No	30086	4%	30011	4%
Yes	Yes	No	No	Yes	2603	0%	2606	0%
Yes	Yes	No	Yes	No	1388	0%	1391	0%
Yes	Yes	No	Yes	Yes	5237	1%	5226	1%
Yes	Yes	Yes	No	No	29785	4%	29790	4%
Yes	Yes	Yes	No	Yes	6276	1%	6274	1%
Yes	Yes	Yes	Yes	No	34261	5%	34238	5%
Yes	Yes	Yes	Yes	Yes	320547	48%	319673	47%

## Missing records in SWC

A teacher may be teaching in a school and yet not recorded in the November School Workforce Census for a number of reasons:

1. The school may have entirely failed to make a SWC return. Where this happened in 2010, 2011, 2012 and 2013, the Database of Teacher Records has been used to impute the missing information from Teachers' Pensions Scheme records. But there is considerable lag between a teacher's arrival at a school and their appearance in the Database of Teacher Records, so – particularly for new teachers – this will not always be possible.
2. The school may have inadvertently failed to record an individual teacher. This may be more likely where that teacher is new to the school payroll and was not included in the previous year's census return.
3. The school may have decided not to provide a return for teachers currently training in their school if they did not consider them to be 'full' employees, against the completion guidance provided by the DfE.

If the missing record problem is significant, particularly in relation to those who are new to the profession or to the school, then we risk seriously understating teacher retention as there will appear to be more cases of a teacher being in the school system, then disappearing.

## Known missing records for employment-based ITT teachers

### Teach First placements in years one and two

Teach First participants are employed by schools in the first two years of their programme and so should appear in SWC. Teach First maintain their own records of participant location and drop-out. From these records we can see that across the cohorts starting the programme in 2010, 2011 and 2012:

- Only 81 per cent can be identified in SWC in the November of year one, whereas Teach First's own records suggest 96 per cent are teaching in schools;
- 78 per cent can be identified in SWC in the November of year two, whereas Teach First's own records suggest 87% are teaching in schools.

### School Direct Salaried

School Direct is a new programme - the salaried route was first available from 2013/14. These School Direct Salaried trainees are employed while training and should appear in SWC during their training year. NCTL's initial analysis of the School Direct salaried route showed very significant numbers of missing records for the 2185 trainees in the 2015 ITTPP achieving QTS at the end of their course. Of these:

- 2183 were allocated a TRN
- 1581 were found in SWC during their training year (2013)
- 1768 were found during their first year of qualified employment in SWC 2014 (1405 appear in both training and first post-QTS year).

## Finding missing records using the date of arrival in school

The SWC contains a field recording the date of arrival in school of the teacher. So, for example, if a teacher appears in SWC 2012 with a date of arrival in school of 01/09/2010 then we should find the same teacher in the same school in SWC 2011 and SWC 2010. Assuming the date of arrival in school is accurate, if we cannot find them in earlier years then we know the records are missing and can write the records back into SWC. We can do this for two different types of SWC records: those for teachers who are still teaching at the school (open records) and those for teachers who left the school within the past 12 months (closed records). Table 6 shows that the number of records found via open contracts varies a little by SWC year; very few additional records are identified through analysis of the arrival date on closed contracts.

There are two possible difficulties with the approach. The first is that teachers may be incorrectly written back for years where they were not actually *employed* teachers in the school:

- This might happen where students are on PGCE training placements
- The date of arrival in school may simply be incorrect
- There may be breaks in service within the period bounded by the date of arrival and current date.

The second possible difficulty is that the probability of recovering the record may be a function of the characteristics of the teacher or the school. In particular, it seems less likely that we would manage to write-back a missing record for high turnover teachers than we would for those with a tendency to remain in schools for longer. It is also possible that an academy resets the date of arrival in school at the point of academy conversion (although our inspection of the data suggests they are only resetting the contract start date and not the arrival date). Thus, all of these adjustments to write back missing records may or may not further distort the presence of records in SWC.

Overall, this approach to recovering missing records materially affects estimates of early career retention since it disproportionately writes back records for teachers who are young and who are recently qualified.

**Table 6 Identifying missing records using date of arrival in school**

	Found in main SWC		Found via arrival date on open contract		Found via arrival date on closed contract	
	N	%	N	%	N	%
2010 SWC	485416	94%	25422	5%	4590	1%
2011 SWC	482978	96%	21267	4%	1252	0%
2012 SWC	488728	96%	17154	3%	1815	0%
2013 SWC	492596	97%	12704	2%	3916	1%

Table 7 summarises the success of the use of arrival date to recover Teach First participant records that are known to be missing. It shows the importance of this write back approach for participants in the first year of the programme, with fewer records written back in year 2. There are still some records missing for each cohort and year.

**Table 7: Success rate in identification of Teach First records in School Workforce Census**

		Cohort starting the programme in:		
		2010	2011	2012
<b>Year 1:</b>	% of starting cohort found in SWC	81%	79%	82%
	% found using open contract start dates	7%	8%	6%
	% found using closed contract start dates	1%	1%	2%
	% found somewhere in SWC	89%	88%	90%
	% of starting cohort that TF record as still teaching	96%	96%	96%
<b>Year 2:</b>	% of starting cohort found in SWC	81%	76%	77%
	% found using open contract start dates	2%	1%	1%
	% found using closed contract start dates	0%	1%	3%
	% found somewhere in SWC	83%	78%	81%
	% of starting cohort that TF record as still teaching	90%	85%	86%



Department  
for Education

© Education Datalab, FFT 2016

**Reference: DFE- RR507**

**ISBN: 978-1-78105-568-7**

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at:

[ann.claytor@education.gsi.gov.uk](mailto:ann.claytor@education.gsi.gov.uk) or [www.education.gov.uk/contactus](http://www.education.gov.uk/contactus)

This document is available for download at [www.gov.uk/government/publications](http://www.gov.uk/government/publications)