TECHNICAL PAPER

Apprentice and trainee statistics:
estimation of contract completion and attrition rates

June 2010

BRIAN HARVEY

NATIONAL CENTRE FOR VOCATIONAL EDUCATION RESEARCH

Australian Government

Department of Education, Employment and Workplace Relations

Contents

Tables

Completion and attrition rates

Introduction

Apprentice and trainee data are reported by the state and territory training authorities to the National Centre for Vocational Education Research (NCVER) on a quarterly basis, beginning at the September quarter of 1994. The set of data submitted that quarter is referred to as the National Apprentice and Trainee Collection 1. The sets of data submitted in subsequent quarters are referred to as Collection 2, Collection 3 and so on. At the time of writing, the set of data being submitted is for the March 2010 quarter and is referred to as Collection 63. The data consist of information about the contracts of training entered into by apprentices and employers.

The demand for information on the proportion of apprenticeships that result in successful completions and the proportion of apprenticeships that result in cancellations or withdrawals (that is, attrition) has grown over recent years. The obvious way to determine these proportions is to track each contract from its commencement and to check whether it has resulted in an outcome or not. If the contract has an outcome, it will be either a completion or attrition (that is, a cancellation or withdrawal). This principle is sound and has been used in the past by considering cohorts of contracts that commence in a given year.

For contracts that commence in a given time period (usually calendar year), the proportions of attrition and completion can be calculated at any subsequent point in time. Clearly, enough time must pass after the commencement period in order for the proportions to approximate the completion and attrition rates of the commencing cohort. Although attrition can occur at any time after commencement, completion implies meeting all contract requirements, and this generally requires a specified time in training.

In addition to having to wait for enough attrition and completion to occur, there is a further time lag involved in reporting these events to NCVER. The effect of these reporting lags is such that the completions that occur in a given quarter might take about a year or more to be fully reported. For attrition, this is closer to two years.

The cohort method mentioned above is therefore restricted to cohorts of contracts that commence far enough back in time, such that the calculated proportions approximate valid rates and that the reporting delays are no longer relevant. Of these two factors, it is the effect of reporting delays that can be reduced. The National Apprentice and Trainee Collection calculates estimates of commencements, completions, cancellations and withdrawals for time periods deemed to be affected by reporting delays. If these estimates are used to calculate proportions, then it should be possible to consider commencements from more recent years than would otherwise be the case.

When considering the more recent commencement years, it is more likely that not enough time has elapsed for the proportions completed or for the proportions of attrition to be considered as 'final' rates. In this situation, the proportions are just the 'rate to date' and can be expected to change appreciably as time passes. If, for a given commencement year, the proportion of contracts still 'in training', or yet to report valid outcomes, is high, compared with previous commencement years, then this indicates that more time needs to pass before the calculated proportions of completion or attrition can be considered as valid estimates of the corresponding rates.

The remainder of this document details the method adopted by NCVER to calculate the proportions of attrition and completions and was first used in the 2009 Apprentice and Trainee Annual publication. The next section describes the method in overview and the following section describes the method in detail.

Description of the method

As mentioned above, NCVER calculates estimates of commencements, completions, cancellations and withdrawals when it is deemed that the corresponding counts are affected by reporting delays. Details on the estimation methodology can be found at <http://www.ncver.edu.au/publications/2267.html>. The key outcome from the estimation process is that each record on the Apprentice and Trainee database has a weight associated with it. The meaning of the weight is best explained by way of an example.

Suppose the Apprentice and Trainee database contains 100 completion records in a given state/territory for a given quarter. Suppose further that it is estimated that 150 completions actually occurred; that is, another 50 completions are yet to be reported. The ratio of actual completions to reported completions is $150/100 = 1.5$. This ratio is the weight that is associated with each of the 100 reported completions in the database. The estimate can be reconstructed from the database by adding the weights associated with each of the 100 records in the database.

Similarly, records on the database that relate to other quarters, states/territories and events (that is, commencements, cancellations etc.) will have specific weights associated with them and estimates can be reconstructed as above. Where the records in the database are associated with time periods deemed to be no longer affected by reporting delays, the actual number of events is estimated by the number of reported events, and so the weight for each of these records is one.

The method applies exactly the same way to completions as it does to attrition. Therefore the following discussion will refer to 'outcomes', with the intent that it can be read as either completions or attrition.

Table 1 illustrates the information in the database from the present quarter back to some past quarter, designated as quarter zero.

Table 1   Reported commencements and outcomes

Commenced		Reported								All
		Quarter of outcomes								Outcomes
Quarter	Number	0	1	2	3	4	5	…	Present	Outcomes
0	X_0	$Y_{0,0}$	$Y_{0,1}$	$Y_{0,2}$	$Y_{0,3}$	$Y_{0,4}$	$Y_{0,5}$		$Y_{0,p}$	$Y_{0,+}$
1	X_1		$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,4}$	$Y_{1,5}$		$Y_{1,p}$	$Y_{1,+}$
2	X_2			$Y_{2,2}$	$Y_{2,3}$	$Y_{2,4}$	$Y_{2,5}$		$Y_{2,p}$	$Y_{2,+}$
3	X_3				$Y_{3,3}$	$Y_{3,4}$	$Y_{3,5}$		$Y_{3,p}$	$Y_{3,+}$
4	X_4					$Y_{4,4}$	$Y_{4,5}$		$Y_{4,p}$	$Y_{4,+}$
5	X_5						$Y_{5,5}$		$Y_{5,p}$	$Y_{5,+}$
…										
Present	X_p								$Y_{p,p}$	$Y_{p,+}$
		$Y_{+,0}$	$Y_{+,1}$	$Y_{+,2}$	$Y_{+,3}$	$Y_{+,4}$	$Y_{+,5}$		$Y_{+,p}$	

In table 1,
  X_i is the number of reported commencements in quarter i
  $Y_{i,j}$ is the number of reported outcomes in quarter j that had a commencement in quarter i
  $Y_{+,j}$ is the number of reported outcomes in quarter j
  $Y_{i,+}$ is the number of reported outcomes that had a commencement in quarter i.

Table 2 illustrates the estimates available for the same period as in table 1. In table 2,

X_i is the estimated number of actual commencements in quarter i

$Y_{+,j}$ is the estimated number of actual outcomes in quarter j.

Table 2   Estimated commencements and outcomes—margins

		Estimated								All
Commenced		Quarter of outcomes								All
Quarter	Number	0	1	2	3	4	5	…	Present	Outcomes
0	X_0									
1	X_1									
2	X_2									
3	X_3									
4	X_4									
5	X_5									
…										
Present	X_p									
		$Y_{+,0}$	$Y_{+,1}$	$Y_{+,2}$	$Y_{+,3}$	$Y_{+,4}$	$Y_{+,5}$		$Y_{+,p}$	

In order to calculate the proportions of outcomes by commencement years, estimates are required for the cells in the body of the table. These estimates are obtained by pro-rating the estimated number of outcomes across the rows of the table in the same proportions as the columns of table 1. The estimate for the cell representing commencement quarter i and termination quarter j is therefore:

$$Y_{i,j} = Y_{+,j} \left(Y_{i,j}/Y_{+,j} \right).$$

This means the cells of table 2 can be calculated, giving table 3.

Table 3   Estimated commencements and outcomes—full table

		Estimated								All
Commenced		Quarter of outcomes								All
Quarter	Number	0	1	2	3	4	5	…	Present	Outcomes
0	X_0	$Y_{0,0}$	$Y_{0,1}$	$Y_{0,2}$	$Y_{0,3}$	$Y_{0,4}$	$Y_{0,5}$		$Y_{0,p}$	$Y_{0,+}$
1	X_1		$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,4}$	$Y_{1,5}$		$Y_{1,p}$	$Y_{1,+}$
2	X_2			$Y_{2,2}$	$Y_{2,3}$	$Y_{2,4}$	$Y_{2,5}$		$Y_{2,p}$	$Y_{2,+}$
3	X_3				$Y_{3,3}$	$Y_{3,4}$	$Y_{3,5}$		$Y_{3,p}$	$Y_{3,+}$
4	X_4					$Y_{4,4}$	$Y_{4,5}$		$Y_{4,p}$	$Y_{4,+}$
5	X_5						$Y_{5,5}$		$Y_{5,p}$	$Y_{5,+}$
…										
Present	X_p								$Y_{p,p}$	$Y_{p,+}$
		$Y_{+,0}$	$Y_{+,1}$	$Y_{+,2}$	$Y_{+,3}$	$Y_{+,4}$	$Y_{+,5}$		$Y_{+,p}$	

For any given commencement year there are now estimates for the number that resulted in an outcome by how many quarters after commencement, as well as overall. The estimates in this table are aggregated into years of commencement (rows); however, the columns of the table have to be slightly altered to represent the number of quarters until an outcome is achieved, instead of the actual quarter of the outcome. This leads to the estimates as shown in table 4.

Table 4   Estimated commencements and outcomes—annual table

| Commenced | | Estimated | | | | | | | | All |
| | | Number of quarters until outcome | | | | | | | | |
Year	Number	0	1	2	3	4	5	…	Present	Outcomes
0	W_0	$Z_{0,0}$	$Z_{0,1}$	$Z_{0,2}$	$Z_{0,3}$	$Z_{0,4}$	$Z_{0,5}$		$Z_{0,p}$	$Z_{0,+}$
1	W_1		$Z_{1,1}$	$Z_{1,2}$	$Z_{1,3}$	$Z_{1,4}$	$Z_{1,5}$		$Z_{1,p}$	$Z_{1,+}$
2	W_2			$Z_{2,2}$	$Z_{2,3}$	$Z_{2,4}$	$Z_{2,5}$		$Z_{2,p}$	$Z_{2,+}$
3	W_3				$Z_{3,3}$	$Z_{3,4}$	$Z_{3,5}$		$Z_{3,p}$	$Z_{3,+}$
4	W_4					$Z_{4,4}$	$Z_{4,5}$		$Z_{4,p}$	$Z_{4,+}$
5	W_5						$Z_{5,5}$		$Z_{5,p}$	$Z_{5,+}$
…										
Present	W_p								$Z_{p,p}$	$Z_{p,+}$

In Table 4, W_i is the sum of the four X_js that are associated with the quarters of the 'ith' year. For example, $W_0 = X_0 + X_1 + X_2 + X_3$.

$Z_{i,j}$ is the sum of all the $Y_{k,l}$ that are associated with commencing in the 'ith' year and resulted in outcome within j quarters of commencing. For example, $Z_{0,0} = Y_{0,0} + Y_{1,1} + Y_{2,2} + Y_{3,3}$.

For commencement year 'i', the estimate of the proportion of contracts that have resulted in outcome is $Z_{i,+} / W_i$. Similarly, the estimate of the proportion of contracts that have resulted in outcome within 'j' quarters of commencing is $(Z_{i,0} + Z_{i,1} + Z_{i,2} + ... + Z_{i,j}) / W_i$.

Practical considerations

The preceding section shows how to calculate estimated outcome proportions for any given commencement year. These estimated proportions can only be treated as meaningful if enough time has passed since commencement to allow for the outcomes to occur. There is no rule which guarantees when enough time has passed, so some judgment is required.

Historically, outcome rates have remained at about the same level, except when the number of commencing contracts being considered is small. If the outcome rate for a given year is much lower than the equivalent rate for previous years, then this is an indication that more time is required for that rate to be treated as a valid estimate.

This indication is reinforced if the proportion of 'unaccounted' contracts is much larger than for the previous years. The proportion of unaccounted contracts is simply one minus the sum of the completion and attrition rates. Unaccounted contracts include those yet to report a valid outcome and those still in training.

Overall, the data quality is adequate for the application of this method. There is a small number of contracts with different values for some data items on their commencing and outcome records. These records were 'cleaned' by ensuring that the values on the commencement record were used for the outcome record. For example, where outcome rates are calculated 'by ANZSCO'[1], it is referring to the ANZSCO recorded on the commencements.

The inclusion of transfers as a component of attrition will become an issue as the use of the category increases. There is an inconsistency between jurisdictions which affects whether a contract needs to transfer to a new contract or not. If an apprentice needs to change employer, some jurisdictions can keep the same contract and just amend the employer field. Other jurisdictions must end the existing contract and begin a new one. The solution to this problem is to 'clean' the

---

[1]   ANZSCO = Australian and New Zealand Standard Classification of Occupations.

data before applying the method. Reliable procedures need to be developed to link the contracts involved in the transfer and create one 'derived contract', which can be used instead of the linked contracts. The method can then be applied as described above.

Glossary of terms

Collection – refers to the National Apprentice and Trainee Collection, a set of data that is submitted quarterly by the state and territory training authorities to NCVER. The data are cumulative from a set point in time (currently data are submitted from 1 July 2002 to the current date). Collections are numbered sequentially from the first collection submitted to NCVER in 1994.

Reporting lag – refers to the delay in time between an event (commencement, completion …) occurring and the reporting of that event to NCVER via the state and territory training authority data submissions.

Commencement – refers to the start of a contract of training. The commencement date is the actual date an apprentice or trainee commenced training under a training contract.

Completion – refers to a contract of training in which all of the prescribed requirements have been met.

Completion rate – refers to the proportion of completions resulting from a given set of commencements.

Attrition – refers to a contract of training which ends before the term of the contract has expired, that is, i.e. cancelled, withdrawn or transferred.

Attrition rate – refers to the proportion of attrition resulting from a given set of commencements.

Outcome – since the method is the same for completion and attrition, this note uses 'outcome' as a generic term. The reader may substitute either 'completion' or 'attrition' for outcome as appropriate.

Rate to date – refers to term used for calculated outcome proportions when the commencement year is too recent to allow enough time for the commencing contracts to have outcomes in sufficient numbers.

Other related terms – see the supporting documents for the latest apprentice and trainee annual publication at <http://www.ncver.edu.au/statistic/21049.html>.