NATIONAL VOCATIONAL EDUCATION
AND TRAINING RESEARCH PROGRAM

TECHNICAL PAPER

Identifying the causal effects of VET qualifications on student outcomes through matching

Thorsten Stromback
CENTRE FOR LABOUR MARKET RESEARCH
CURTIN UNIVERSITY

NCVER Building Researcher Capacity
Fellowship Program 2011 participant

NCVER

Australian Government
Department of Industry, Innovation
Science, Research and Tertiary Education

NCVER

Australian Government

Department of Industry, Innovation
Science, Research and Tertiary Education

Identifying the causal effects of VET qualifications on student outcomes through matching

Thorsten Stromback

Centre for Labour Market Research, Curtin University

Participant in the NCVER Fellowship Program 2011

NATIONAL VOCATIONAL EDUCATION
AND TRAINING RESEARCH PROGRAM

TECHNICAL PAPER

As part of the National Centre for Vocational Education Research (NCVER) Building Researcher Capacity Scheme, a fellowship program has been created to encourage researchers to use NCVER datasets to improve our understanding of tertiary education. The fellowships also provide participants with an opportunity to have their research peer-reviewed and published by NCVER.

For more information see: <http://www.ncver.edu.au/research/opportunities.html#NCVER_Fellowships>.

This document should be attributed as Stromback, T 2012, *Identifying the causal effects of VET qualifications on student outcomes through matching*, NCVER, Adelaide.

ISBN 978 1 922056 30 6
TD/TNC 110.02

Published by NCVER, ABN 87 007 967 311

Level 11, 33 King William Street, Adelaide SA 5000
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

P +61 8 8230 8400 **F** +61 8 8212 3436 **E** ncver@ncver.edu.au **W** <http://www.ncver.edu.au>

About the research

Identifying the causal effects of VET qualifications on student outcomes through matching

Thorsten Stromback, Centre for Labour Market Research, Curtin University

This paper's primary focus is methodological. Its purpose is to show how matching methods can be used to estimate the effects of a treatment, such as completion of a VET qualification.

The experimental sciences have a huge advantage over the social sciences. Experiments can be carefully designed to isolate the effect of a treatment. Such an approach is rare in the social sciences, where typically a group of people is observed, some having been subjected to a 'treatment' and others not. Given such data, a simple comparison of the outcomes between the treated group and the untreated group can be quite misleading because the characteristics of the two groups can be quite different. To overcome this, multivariate statistical models can be built. These models can become very complicated and incorporate assumptions which may or may not be reasonable. The complexity of the models is also limited by the number of observations.

Matching methods offers an alternative to multivariate modelling. This approach is intuitively attractive and involves comparing the outcomes of the treated group with a comparison group. The statistical rigour is obtained from the construction of the comparison group, such that for each member of the treatment group there is an individual in the comparison group with very similar characteristics. The method by which the comparison group is constructed involves estimating a probability that an individual is in the treatment group on the basis of background characteristics. An individual in the treatment group is matched with an individual not treated on the basis of having the same probability of being in the treatment group.

In some cases, there is a background variable that is so important that a match on this characteristic is also required. For example, if the outcome variable is *being employed*, then it is critical that the treated individuals have the same employment status as their untreated peers, for the simple reason that the best predictor of being employed after training is being employed before training.

Using Student Outcomes Survey data to look at the effect of qualifications on outcomes after training, the matching techniques are illustrated. The main point to emerge was that, for vocational educational education and training (VET) graduates, higher-level qualifications, on average, increase earnings, improve employment outcomes and are considered more relevant to jobs than lower-level qualifications.

With regard to module completers, however, it is shown that the method cannot be used to estimate: the relative effects of module completions at different levels; or, the effects of completing a full qualification, as opposed to only completing some modules. This is because the background characteristics do not provide a good prediction of who will complete.

This research was funded through the NCVER fellowship program, which encourages researchers to use NCVER datasets.

Tom Karmel
Managing Director, NCVER

Contents

Tables and figures

Tables

Figures

Introduction

The effect of vocational education and training (VET) on student outcomes has been a major issue in VET research and many methods have been applied to many different datasets. This paper contributes to this research agenda by demonstrating how matching methods can be used to estimate the effects of VET qualifications. Based on data from the 2009 Student Outcomes Survey (SOS), it estimates the relative effects of different levels of VET qualifications on students' outcomes. In relation to graduates it is found that higher-level qualifications lead to better outcomes. With regard to module completers, however, it is shown that the method cannot be used to estimate: the relative effects of module completions at different levels; or, the effects of completing a full qualification, as opposed to only completing some modules. Credible estimates of these effects depend on richer data and methods with the ability to take account of unobservable effects.

Background

A significant part of the National Centre for Vocational Education Research's (NCVER) research agenda is concerned with the causal effects of programs and policies. To date, two main methods have been used: descriptive analysis of the data and multivariate statistical models. In a descriptive analysis the outcomes for one group are compared with the outcomes for other groups. Based on this comparison, it is often implied that the observed difference is an estimate of the causal effect of the variable(s) that define the groups. Multivariate analysis using standard regression models has a greater claim for being able to identify causal effects. However, standard models are based on strong assumptions that are often not satisfied. But relaxing these assumptions gives rise to very complex models. The first approach may be seen as unduly simplistic, while the second approach demands extensive data and analysis and may risk the overstatement of impacts.

Matching methods can be seen as combining the best of these two approaches. The analysis looks very much like a descriptive analysis; the researcher can see what they are doing and, because of the strong intuitive appeal of the methods, the technical issues are less daunting. At the same time, matching methods, like multivariate statistical models, can estimate causal effects under assumptions that are plausible in many cases. An additional advantage is that matching allows for heterogeneous effects in a general way. This is important when there is reason to believe that the effect of programs or policies may differ between individuals.

The objective of this paper is to illustrate how matching can be used to estimate the relative effect of different levels of VET qualification. Hitherto there has been only one application of matching to Australian VET (Stromback 2010). In view of the extensive use of matching in the international literature, there is much scope for an increased use of this method in VET research.

Overview

The Student Outcomes Survey (SOS) is a major source of information about the outcomes of VET students. Being a survey of participants only, it has nothing to say about the effect of completing a VET versus not completing a VET qualification. But it is informative about the relative effects of one qualification versus another and includes some variables that influence the choice of qualification and outcomes. Strictly speaking, the identification of a causal effect requires that the researchers observe all the variables that influence the choice of qualification and outcome, but this condition is rarely satisfied in practice.

Using the Student Outcomes Survey data the paper uses matching to estimate two causal effects of the relative effect of one VET qualification versus another VET qualification: the average relative effect of a qualification for those who undertook this qualification and the average relative effect of a qualification for those who did not undertake this treatment but the other qualification. In short, these effects are the average effect on the treated and non-treated, respectively.

More specifically, and making use of the treatment-effect paradigm, each VET qualification is regarded as one of M treatments and the objective of the analysis is to estimate the average effect of treatment i relative to treatment j for individuals who undertake treatment i and j, respectively, i,j = 1,…M. In plain language it is proposed to estimate, for example, the average effect of undertaking a certificate III qualification rather than a certificate II qualification for individuals in both groups. In the first case the estimate is informative about the incremental effect of the certificate III qualification for those who actually undertake it. In the second case the estimate tells us what certificate II students would stand to gain if they had done a certificate III qualification instead. Lack of overlap may prevent the estimation of all pairwise effects, that is, diploma students may be too different from certificate I students, but that is not a significant limitation.

Methodology

The basic ideas and concepts of current approaches to the estimation of causal treatment effects stem from the statistical analysis of randomised experiments and potential outcomes.[1]

Using the potential outcome concept (Rubin 1974), if each person participates in exactly one of the M treatments, there is a set of potential outcomes for each person:

$Y^0, Y^1, Y^2, \ldots, Y^{M-1}$.

where Y^0 is the outcome when the individual is assigned to treatment 0, Y^1 the outcome when the individual is assigned to treatment 1 and so forth. Before the event of participation, each of these potential outcomes is latent. After the event, the only outcome observed is the outcome of the treatment the person actually participated in. The other potential outcomes are counterfactual and unobservable by definition. Since only one of the potential outcomes is observed, the joint distribution of $Y^0, Y^1, Y^2, \ldots, Y^{M-1}$ is not identified; at most the marginal properties of their marginal distributions can be identified.

What can be identified under certain conditions are certain average treatment effects. The average treatment effect (ATE) is defined as:

$E[Y_i - Y_j]$,

the expected difference in the outcomes between participating in treatment i and treatment j for a person drawn at random from the population of individuals participating in either i or j.

Analogously, the average treatment effect on the treated (ATT)

$E[Y_i - Y_j \mid D=i]$

is the expected outcome difference for a person drawn at random from the sub-population of participants in treatment i. This parameter is commonly held to be the most relevant parameter, particularly when the participation is voluntary. Nevertheless, the average treatment effect of treatment i on those who participated in another treatment j (ANT)

$E[Y_i - Y_j \mid D=j]$

might also be relevant if one wished to induce individuals currently receiving one treatment to take up another treatment.

Strategies for identifying any of these average treatment effects all rely on comparing the observed outcomes of one group of individuals with the observed outcomes of another group. If the assignment to treatment is random, the potential outcomes are statistically independent of the treatment. This ensures that any differences between the treatment groups are due to chance alone and not systematic. Consequently, the observed outcome for participants in one treatment has the same expected outcomes as the potential outcomes for participants in another treatment.

The method of matching mimics the random assignment process by forming comparison groups that are as similar as possible to the groups that are treated. Matching has become a widely used method for estimating causal treatment effects in business, social and educational research (Caliendo & Kopeining 2008). The underlying idea is that if two individuals have very similar characteristics (X)

[1] Recent overviews of current approaches can be found in Dehijia and Wahba (2002), Frolich (2002), Caliendo and Hujer (2006) and Imbens and Wooldridge (2008).

their potential outcomes should also be similar. If one person takes part in treatment i and the other in treatment j and there are many such pairs, then the difference in observed outcomes can be used as an estimate of the relative effect of treatment i versus treatment j. For these estimates to be consistent, the individuals within each pair must be identical with relation to all confounding variables, all variables that influenced the selection into treatment and the potential outcomes. This implies that, conditional on these confounding variables, the probability of participating in a particular treatment is independent of the potential outcomes. This assumption is known as the ignorable treatment assignment or conditional independence assumption.

As implied by the above, matching depends on forming comparison groups that are similar with respect to the confounding variables: for each person in treatment i we have to find a counterpart in treatment j. This common support requirement restricts the scope of matching estimators to the sub-population who meets this condition. This makes intuitive sense. Matching works by comparing comparable individuals. If two groups are not comparable, the best we can do is to estimate a treatment effect for the subset of individuals within these groups that are comparable.

When the set of confounding variables is large, finding a sufficient number of similar individuals is difficult. In the case of binary treatments Rosenbaum and Rubin (1983) showed that the independence of selection into treatment and the potential outcomes conditional on the confounding variables also implies independence conditional on the probability of participating in a treatment. This probability they called the propensity score. This result has been generalised to the multiple treatment case, where the appropriate propensity score is the probability of participating in treatment i for an individual with characteristics X who participates either in treatment i or j (Imbens 2004; Lechner 2001). The intuition is that the propensity score balances the distribution of the confounding variables in the two treatment groups. Propensity-score matching circumvents the dimensionality problem, permitting matching to be performed with respect to the one-dimensional propensity score.

While matching on the propensity score balances the distribution of the confounding variables in treatment groups, it may not be the most precise estimator in finite samples: the components of X might affect the propensity score and the conditional expectation function to different degrees. In the Student Outcomes Survey it is well understood that one outcome, post-training employment, is highly dependent on one particular confounding variable, pre-training employment status. In this case, observations with similar propensity scores may not be very similar with respect to this particular variable; its main influence is not on the probability of a particular treatment but on one particular outcome. To achieve balancing on the relevant variables in this case, matching can be carried out on both the propensity score and this confounding variable. This method is commonly called augmented propensity-score matching (Bryson, Dorsett & Purdon 2002; Lechner 2002).

Matching vs regression

The most common method for estimating causal effects is regression. In a regression framework an outcome variable is regressed on the treatment status as well as on all the other confounding variables that determine the outcome. The identification of a causal treatment effect in this framework also relies on the unconfoundedness assumption: that, conditional on the confounding variables, the assignment to treatment is independent of the outcome.

There are, nevertheless, three key differences between matching and regression that motivate the use of matching (Caliendo & Hujer 2006).

The first difference is that matching avoids the functional form (linearity) assumption, both in the estimation of the structural parameters and in the implicit assumption that linear conditioning eliminates any selection bias.

Secondly, whereas matching explicitly relies on the common support assumption, ensuring that comparisons between treated and controls are only made over the region of common support, regression estimators do not. Regression yields estimates even in the absence of similar comparison units, since the linear functional form fills in for the missing data. If common support is lacking, regression will estimate the model for the untreated in the regions where the untreated observation lies and then project this model into the region where the treated units lie.

Thirdly, matching leaves the individual causal effects unrestricted and allows for individual heterogeneity in the population. In a regression model, on the other hand, the casual effect is assumed constant for all units.

The first and last of these differences between matching and regression fade as higher-order and interaction terms are included in a regression model. However, it is not common to do so, for mainly two reasons. There are no simple rules for which additional terms should be included, and as more terms are included, the interpretation of regression estimates becomes more and more complicated.

The Student Outcomes Survey

NCVER's Student Outcomes Survey is an annual, nationally stratified, randomly selected sample of students who completed their vocational education and training during a year. Completing students are classified into one of two groups: graduates and module completers, depending on whether they were awarded a qualification or not.

The Student Outcomes Survey has large sample sizes of between 81 000 (plus 'top ups' by individual providers) and 300 000 every alternate year. This allows for national and state reporting to be carried out every year, with institute reporting in alternate years (the large sample years).

Against the two advantages, a limitation of this survey is the low survey response rate, at around 40% for graduates and 30% for module completers. The 2009 survey, a large sample year, used in this report contains information about 52 859 graduates and 22 463 module completers. The completion status of the remainder cannot be determined. The low response rate means that the survey is not representative of the population of interest. In the 2008 report, a post-survey analysis of non-respondents found that they differed from respondents in several ways. A particular concern is that those with better outcomes are more likely to respond, leading to overestimates of the benefit of VET. However, this possibility does not preclude the analysis of relative effects, how the outcomes vary with the level of qualification, or comparisons between graduates and module completers. It remains the case, however, that the high level of non-response means that the respondents are not a random sample of the population of VET completers. The survey data include sample weights, which can be used to partially correct for this. These weights are derived from post-survey stratification over a small number of variables. Using these weights to compute matching estimates, which are conditional on a large number of covariates, does not necessarily improve these estimates. Thus, and in line with the vast majority of empirical studies, this paper uses unweighted sample data.

The descriptive summaries of the Student Outcomes Survey contain a wealth of information about the characteristics of individuals who participate in VET and the outcomes from this participation. There are also a number of more extensive descriptive studies. Stanwick (2005, 2006) uses the Student Outcomes Survey to analyse the employment and further study outcomes of VET students. In the case of higher VET qualifications he notes that, while the employment outcomes are good, the higher VET qualifications are also an important pathway to university studies. For the lower-level qualifications the employment outcomes are less satisfactory, as is the transition to higher-level qualifications. Another example is Karmel, Mlotkowski and Awodeyi (2007), who analyse the match between the training and the subsequent occupation. They report that this match is not as close as one might expect, suggesting there is a generic component to most VET courses.

In multivariate statistical analyses of Student Outcomes Survey data, the typical method has been to regress an outcome variable on a set of independent variables, where the interest is focused on a subset of these variables. An example is Karmel and Nguyen (2007), who use the survey to study the effect of VET qualifications on earnings. Using a regression model to estimate the effect of the highest qualification or the latest qualification, they find that only higher-level VET qualifications have a statistically significant positive effect on earnings.

Empirical method

To implement the matching method using the Student Outcomes Survey data, completing students are classified according to the level of qualification completed. The qualifications are grouped into five categories: diploma and above and certificates IV, III, II and I. Qualifications below certificate I level are not included. More than one-half of the module completers undertook below certificate I level qualifications (other, non-award or unknown), which reduces the number of module completers used in the analysis to about 10 000 individuals. This means that the cell sizes for module completers, that is, the number of cases at each qualification level, can become quite small. The smallest cell for module completers contains only about 800 cases. This reinforces the concern about the representativeness of the sample of module completers in particular. Missing values for one or more of the independent variables further reduce the number of cases included in the analysis. This reduction is, however, quite small. For most of the variables used in the analysis the proportion with missing values is less than 5%.

For each adjacent pairs of qualifications we first estimate the propensity score, the probability of completing qualification i, given that a person completes either i or j ($p^{i|i,j}$). The propensity score is estimated from a binary probability model using only the observation on individuals who did either i or j. The independent variables is the set X, which are all factors which influence both the selection into qualifications and the potential outcomes from obtaining these qualifications.

The pre-training or confounding variables represent factors that influence the choice of qualification and the outcomes of doing particular qualifications. The post-training or outcome variables are indicators of the outcome from completing a qualification.

Pre-training variables

Personal characteristics

- sex
- age
- indigenous status
- country of birth
- speak language other than english at home
- proficiency in spoken english
- disability status
- level of schooling
- level of post-school qualification
- prior experience and skills relevant to training.

Prior employment status

- employed full-time
- employed part-time
- unemployed.

Area variable

- accessibility/remoteness index Australia (ARIA).

Post-training outcome variables

Post-training status

- employed full-time
- employed part-time
- unemployed
- enrolled in further study.

Employed individuals only

- weekly earnings (if employed full-time)
- job-related benefits of undertaking training
- relevance of training to job after training.

Given the estimated propensity score, the average treatment effects are computed by matching on the propensity score and, for some outcome variables, the pre-treatment employment status.

The ATT is estimated as

$ATT = E(Y^i \mid p^{i|i,j}, D=i) - E[E(Y^j \mid X, D=j) \mid D=i]$

where the first term is estimated from the group treated with i and the second term from the mean outcomes of the matched comparison group receiving treatment j. For the ATT the outer expectation is taken over the distribution of the propensity score in the population that received treatment i. For the ANT the expectation is taken over the distribution that received treatment j.

$ANT = E(Y^i \mid p^{i|i,j}, D=i) - E[E(Y^j \mid X, D=j) \mid D=j]$

The estimation was carried out using the Stata module written by Leuven and Sianesi (2003).

Matching can be carried out using a number of different methods. Here we use the most straightforward method of nearest neighbour matching with replacement. This method always gives a close match between a treated and their control. The disadvantage is that the same untreated observation may be used as the control for several treated. This means that much of the information contained in the non-treated sample is not used. However, there is little to guide the choice of matching method, and alternative methods also have their disadvantages. In practice it seems as if the choice of method makes little difference to the results (Smith & Todd 2005).

The output from the analysis is reported in a series of tables, one separate table for each of the outcome indicators, using the following format.

The average effects of completing qualification i vs qualification j for individuals undertaking either qualification i or j					
	Diploma and above	**Certificate IV**	**Certificate III**	**Certificate I/II**	**Other**
Diploma and above	-	x	x	x	x
Certificate IV	z	-	x	x	
Certificate III	z	z	-	x	x
Certificate I/II	z	z	z	-	x
Other	z	z	z	z	-

- Each entry in the table is an estimate of the effect of undertaking qualification i versus qualification j.
- The above main diagonal entries (x) give the effects of undertaking qualification i versus qualification j for those who undertook qualification i – the average relative treatment effect of the treated (ATT).
- The below main diagonal entries (z) give the effects of undertaking qualification j versus qualification i for those who undertook qualification i – the average relative treatment effect of the non-treated (ANT).
- ATT tells us what the treated have gained from undertaking a higher rather than a lower qualification.
- ANT tells us what individuals who have done a lower qualification would have gained had they done a higher qualification.

In interpreting the results it is important to bear in mind that we seek to estimate the relative effects of the most recent qualification. As shown in table 1, graduates who completed a diploma and above qualification were particularly well qualified before they started their training, with one-third already having a qualification at that level. But those graduating with a lower-level qualification were also well qualified: about 12—15% of those graduating with certificate I—III qualifications already had a diploma and above qualification before they started their training. This means that the observed outcomes are strongly influenced by previous qualifications, in addition to the latest qualification. By the same token the relative effect of undertaking the latest qualification depends on the level of the previous highest level of qualification. This makes matching a particularly appropriate method for estimating the relative effects, since matching does not restrict the individual-level effects to be the same. Of course matching still estimates an average effect; what this average effect is depends on the distribution of the pre-training qualifications of the group for which the estimate is made.

Table 1 Distribution of previous highest post-school qualifications by recent qualification undertaken (%)

Recent qualification	Previous qualification							
	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I	Other	No post-school qualification	**Total**
Diploma and above	32.58	17.97	15.51	5.79	0.79	5.75	21.61	**100.00**
Certificate IV	35.98	11.10	22.17	4.76	0.64	7.07	18.27	**100.00**
Certificate III	15.33	4.29	14.13	13.38	2.46	10.79	39.62	**100.00**
Certificate II	12.99	3.58	9.34	9.39	6.06	9.68	48.94	**100.00**
Certificate I	12.65	2.49	6.00	5.82	8.60	9.44	54.99	**100.00**
Total	**20.92**	**7.16**	**14.32**	**9.40**	**3.04**	**9.08**	**36.07**	**100.00**

As previously explained, the identification of these relative treatment effects depends on the selection of the level of VET courses being independent of the outcome conditional on the observed covariates. In turn, this condition requires that we can observe all the variables that determine both the selection to treatments and the outcomes.

The a priori reason for why this assumption might hold is that education is a sequential process through which students progress in a linear fashion from lower to higher levels. If this is the case, the primary determinant of the choice of the level of study is the previous highest level. Of course, the VET system is not very tightly structured, but it is still the case that students' own preferences, and their selection and admission to different courses, are highly dependent on their previous educational attainment and qualifications. In other words, the main influence on students' choices and admission to different levels of VET — the selection into treatments — are observable in the dataset used: the level of schooling and post-school qualifications. VET choices and admission are also influenced by prior experience and skills relevant to training, which are also observable in the dataset. The influences of these variables on the VET choices have been demonstrated in a large body of empirical research (Long, Carpenter & Hayden 1999; Fullerton 2001; Ball & Lamb 2001).

The only systematic factor missing from the analysis are socioeconomic characteristics. Like prior education, the important role of socioeconomic factors is well established from previous empirical research. Socioeconomic factors have been shown to be particularly important influences on young people's academic study path, schooling and higher education, but they have a smaller impact on the choice of the level of vocational education. The most recent evidence on this issue indicates that students with a high socioeconomic status have lower participation in lower-level VET courses (certificates I and II) than other students. On the other hand, participation in higher-level courses is not strongly associated with socioeconomic status (Teese & Walstab 2008).

A priori reasoning can at best establish that most of the relevant factors are taken into account, but this does not preclude there being other, omitted and/or unobservable, factors that influence the selection or outcomes. Ability, for example, is likely to influence both educational choices and outcomes. In the VET system, aptitude may be just as important as ability, but is even more difficult to measure. For that reason it may be desirable to assess the plausibility of the conditional independence assumption. Several approaches have been suggested in the literature: estimating the effect on a variable that is known to be unaffected by the treatment; estimating the difference in outcomes using multiple control groups; and estimating the sensitivity of the result to a hypothetical unobserved variable (Imbens 2004; Rosenbaum 1987; Rosenbaum & Rubin 1983; Rosenbaum 2002). The first two have limited applicability and depend on additional data. The last is a computer-intensive method that might be applied to a few estimated treatment effects, but not to the large number of effects estimated in this paper.

Results

Assessing the quality of matching

The role of matching is to ensure that like is compared with like. In more technical language, matching should balance the characteristics across the treatment and matched comparison group. To assess the extent to which matching on the propensity scores achieves this, the following four steps are informative:

- determination of the predictive ability of the treatment assignment model (Smith & Todd 2005)
- a comparison of the distribution of the estimated propensity scores of the treated and controls (Bryson, Dorsett & Purdon 2002)
- a comparison of equality of the means of the confounding variables before and after matching (Rosenbaum & Rubin 1983)
- an assessment of the reduction in standardised bias (difference in means standardised by the variance) due to the matching (Lechner 1999; Caliendo, Hujo & Thomsen 2007).

Since a very large number of results are presented in this report, the analysis of the matching quality is restricted to two cases only. In the first case, students with a diploma and above qualification are matched with certificate III graduates and, in the second case, those with a diploma and above qualification are matched with certificate I graduates. A comparison between the two cases also serves to highlight how the differences between the groups affect the matching quality.

As regards the first step, it is not necessary for the variables included in the treatment assignment model to be good predictors of the treatment status: what matters is that matching on the estimated propensity score results in good balance. However, if the included variables are poor predictors of treatment status, it is unlikely that matching on the resulting propensity score will result in a good balance. How well these variables predict treatment status is also informative about the validity of the conditional independence assumption. For these reasons it can be useful to examine the estimates in some detail. Table A1 in the appendix shows that most of the included covariates have a statistically significant impact on the choice of qualification levels. As expected, previous schooling and qualifications are the two factors that exert the greatest influence. Thus the probability of doing a diploma is much higher for those with Year 12 schooling and a certificate IV than for individuals with lower levels of education and training. Comparing the two sets of estimates in the table we note that that the predictive power of the model (as measured by pseudo R2) is quite low for the two more closely related qualifications (diploma vs certificate III) compared with the more distant (diploma vs certificate I). Again, this is as expected; it is more difficult to predict the choice between closely related alternatives.

The second step is not a check on the matching process as such, but is informative about the prospect of obtaining good matches. The densities of the propensity scores of our two example cases are given in figures 1 and 2. In the first case, when certificate III students are the non-treated, we note that the propensity scores of the treated (diploma and above) are evenly distributed over the whole range from 0 to 1. In contrast, most of the non-treated have a low propensity score. The difference in the distribution of the propensity scores is much larger when certificate I graduates are the non-treated. In this case almost all the treated have high scores, while the scores of the non-treated are distributed more uniformly over the whole interval. Notwithstanding these differences, the distributions of the treated and non-treated overlap, meaning that there is close match for all the treated among the non-treated. However, the differences in the distributions mean that the same non-treated observations have to be used as matches for many of the treated. This is most pronounced for certificate I graduates. Only a small proportion of these graduates can be used as matches for the high-score diploma and above graduates.

Figure 1 Distributions of propensity scores

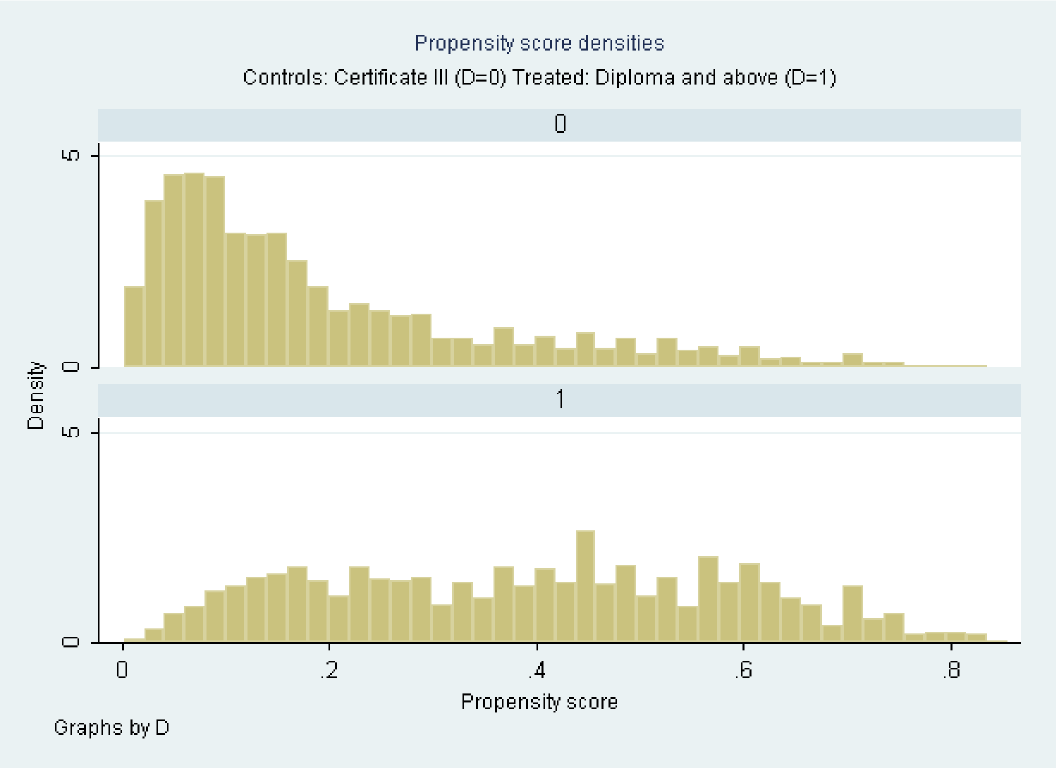
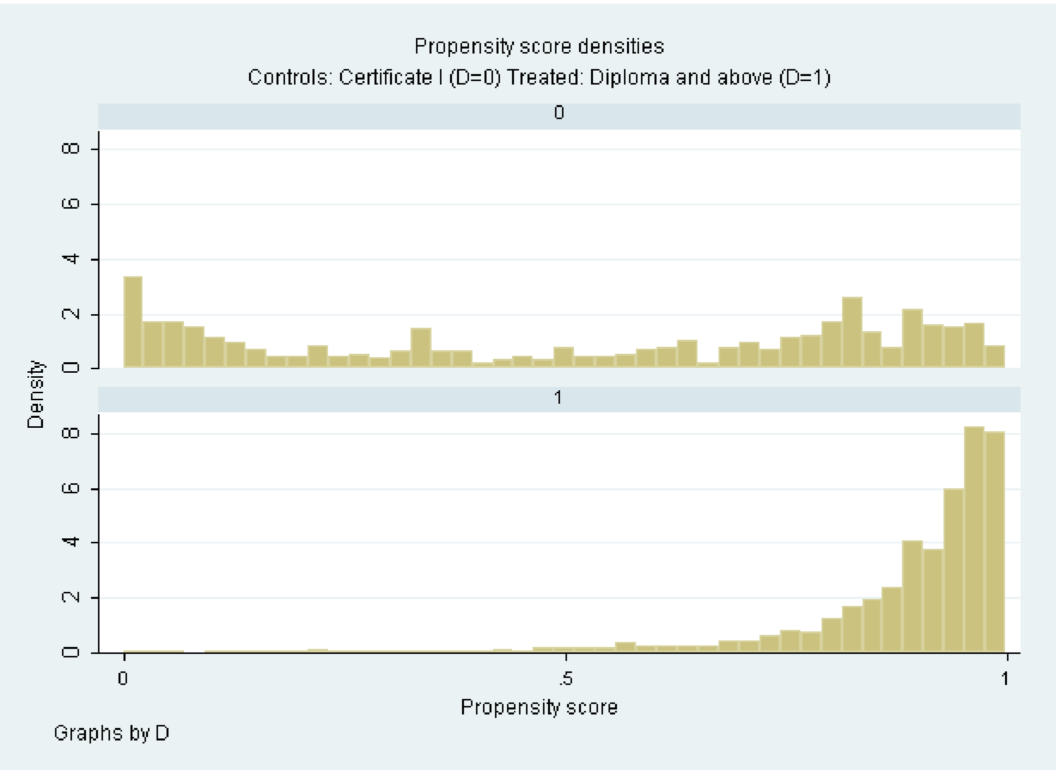


Figure 2 Distributions of propensity scores



The third and fourth checks show that the means of the confounding variables are generally different from the treated and control samples before matching (table A2). After matching, the mean differences are generally very small, and in no case is the difference in means statistically different from zero, according to the t-tests. That matching balances the treated and controls is also evident. Before matching, the standardised bias (difference in means) is generally large, but in many cases the matching reduces the bias by 90% or more. The details in table A2 are summarised in tables 2 and 3. Taking all the confounding variables the median absolute bias is reduced from 12.9% to 1.5% in the diploma versus certificate III case.

Table 2 Summary of the distribution of the standardised absolute bias of covariates before and after matching: diploma vs certificate III

	Percentiles	
	Before matching	After matching
10%	3.02	0.23
25%	7.07	0.59
50%	12.85	1.49
75%	23.49	2.59
90%	35.21	3.39

Table 3 Summary of the distribution of the standardised absolute bias of covariates before and after matching: diploma vs certificate I

	Percentiles	
	Before matching	After matching
10%	3.22	0.10
25%	9.28	1.38
50%	20.46	3.78
75%	29.51	7.60
90%	42.77	12.85

Matching estimates of the relative effects of qualifications on outcomes

The results in the previous section established that matching on the propensity score results in well-balanced treatment and control groups. In this section we report the estimates of the relative effects of qualifications for graduates for each of the outcome indicators: weekly earnings, employment, further study and the benefits and relevance of training.

Weekly earnings

In common with many other datasets, weekly earnings in the Student Outcomes Survey are recorded in bands of about 50 dollars. Since it is cumbersome to work with an outcome variable that takes on a large number of discrete levels, earnings have been recoded to their mid-point values. This approach has been almost universal in the large empirical literature on the determinants of earnings, motivated by the presumption that the measurement errors introduced are random and hence have no effect on the parameters of interest. The data also have the further limitation of having no measure of hours worked; hence, weekly earnings are only comparable for individuals in full-time employment.

Table 4 shows how average weekly earnings vary with the level of qualification attained. There is little difference between the highest two levels but much larger differences between the two highest levels and the lower levels. Thus certificate IV and above graduates earn about 300 dollars per week more than the certificate I graduates.

Table 4 Average weekly earnings by qualification: graduates in full-time employment ($)

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
1026	1044	863	817	747

These comparisons of the means are partly a result of the difference in the pre-training characteristics of the respective groups. Using matching to control for these differences resulted in the estimates given in table 5.

Table 5 Estimates of the effect of qualification i vs qualification j on weekly earnings for individuals who have competed either qualification i or j: graduates in full-time employment after training

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		20 (1.23)	143 (9.30)	105 (5.13)	89 (2.22)
Certificate IV	26 (1.58)		126 (8.61)	85 (4.55)	86 (2.38)
Certificate III	90 (4.41)	87 (5.00)		60 (3.52)	74 (2.11)
Certificate II	128 (4.68)	118 (5.26)	12 (0.88)		19 (0.71)
Certificate I	160 (3.03)	106 (2.48)	52 (2.00)	-5 (0.18)	

The above main diagonal entries give the estimated effects of undertaking qualification i versus qualification j for those who undertook qualification i — the average (relative) treatment effect on the treated.

In the first row the treated are diploma and above graduates whose earnings are compared with what their earnings would have been had they undertaken a lower-level qualification. According to the estimates, the effect of the diploma is to increase their weekly earnings by $20 per week compared with what they would have been had they done a certificate IV instead. The other estimates in the first row give the corresponding figures for the other alternative qualifications. Generally speaking, these estimates are much smaller than the raw difference between qualifications in table 4. This is because diploma graduates have characteristics that predispose them to earn more than graduates with other qualifications. We note, however, that the relative effect of a diploma is smallest when a certificate I is the alternative qualification. This estimate is improbably low and possibly due to imbalanced matching.

The below main diagonal entries give the effects of undertaking qualification j versus qualification i for those who undertook qualification i — the average relative treatment effect on the non-treated. The first column gives the corresponding estimates for the effect of a diploma qualification for those who undertook a lower-level qualification. If certificate IV graduates had done a diploma, it is estimated that they would have earned $26 more per week. This estimate is not very different from the corresponding estimate for those who actually did a diploma. The other estimates in the first column are also close to the estimates in the first row, with the exception of the estimate for certificate I graduates. We also note that, on the whole, the estimates of the ATT are not substantially different from the ANT estimates. In other words, the relative effects are more or less the same, irrespective of which qualification they actually did.

Taking the estimates as a whole there are only small relative effects of diploma vs certificate IV and certificate II vs certificate I. Thus for practical purposes the results suggest a three-level tier of qualifications, but the position of the middle level (certificate III) is difficult to pin down. The effect of undertaking the highest level qualifications is to increase weekly earnings by about $100 per week compared with the two lower levels, while a certificate III yields a smaller increase.

Employment outcomes

The employment outcomes are categorised as being in full-time employment, part-time employment and being unemployed. For each of these outcomes we estimate the average treatment effect of undertaking one qualification relative to the other qualifications. The residual outcome, not in the labour force, is not considered explicitly but is implied by the other three.

Table 6 Proportion in full-time employment: all graduates

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.495	0.537	0.539	0.355	0.250

As regards full-time employment, there is sharp division between the three higher-level qualifications (diploma and above, certificates IV and III) and the two lower (certificates I and II). About 50% of graduates who undertook a higher-level qualification are in full-time employment. For those with lower-level qualifications the corresponding figure is about 30% (table 6). The principal implication of the estimates of the treatment effects in table 7 is that most of this difference is not a causal effect but is due to the pre-training characteristics of the respective groups. The above diagonal estimates of the ATTs show that the relative effect of the higher-level qualifications vis-à-vis the lower-level qualifications on the probability of being in full-time employment are positive but modest. In the fourth column, where the higher levels are compared with a certificate II, the estimates are particularly small. When the comparison is with certificate I, the estimated effect is larger, but this should be seen in the light of the very large raw difference (about 25 percentage points) between certificate I graduates and higher-level graduates. The ANT estimates tell essentially the same story. Had the lower-level graduates undertaken a higher-level qualification, their probability of being in full-time employment would not have increased by much.

Table 7 Estimates of the effect of qualification i vs qualification j on the probability of being in full-time employment for individuals who have competed either qualification i or j: all graduates

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		-0.044 (2.16)	-0.034 (1.74)	0.012 (0.54)	0.072 (1.55)
Certificate IV	0.031 (1.65)		0.021 (1.15)	0.018 (0.92)	0.077 (2.69)
Certificate III	-0.041 (2.10)	-0.032 (1.92)		0.093 (6.17)	0.159 (7.11)
Certificate II	0.032 (0.66)	0.041 (1.42)	0.079 (4.51)		0.064 (3.32)
Certificate I	0.008 (0.09)	0.093 (2.08)	0.095 (4.11)	0.057 (2.68)	

In the case of part-time employment there are no large differences between the qualification levels (tables 8 and 9). The estimates of the relative effects of qualifications are also small and only one estimate is significantly different from zero at the 5% level. Thus, the implication of table 9 can be stated very succinctly: which qualification you do has no real effect on the probability of being in part-time employment after training.

Table 8 Proportion in part-time employment: all graduates

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.330	0.313	0.290	0.350	0.289

Table 9 Estimates of the effect of qualification i vs qualification j on the probability of being in part-time employment for individuals who have competed either qualification i or j: all graduates

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		0.031 (1.69)	0.013 (0.72)	0.008 (0.41)	0.067 (1.52)
Certificate IV	-.019 (1.13)		-0.019 (1.17)	0.022 (1.31)	0.004 (0.15)
Certificate III	0.019 (1.07)	0.031 (1.99)		-0.029 (2.14)	-0.039 (1.78)
Certificate II	-0.003 (0.06)	0.029 (1.10)	-0.028 (1.72)		-0.002 (0.11)
Certificate I	0.006 (0.08)	0.072 (1.76)	0.018 (0.83)	0.017 (0.88)	

In the case of unemployment the estimates are negative, meaning that undertaking a higher qualification reduces the chances of being unemployed (table 11). Furthermore, and in contrast to what was observed for full-time employment, the impact of a higher qualification is quite large relative to the mean differences between qualification levels (table 10). The largest estimate, the effect of a diploma versus certificate I, is to reduce the probability of being unemployed by 8.4 percentage points. This is of the same order of magnitude as the difference in the proportion of the unemployed between the two groups. Other estimates are smaller and not always significantly different from zero. Nevertheless, taking the estimates as whole, higher-level qualifications reduce the chance of being unemployed after training.

Table 10 Proportion unemployed: all graduates

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.102	0.070	0.089	0.146	0.199

Table 11 Estimates of the effect of qualification i vs qualification j on the probability of being unemployed for individuals who have competed either qualification i or j: all graduates

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		0.031 (2.54)	0.020 (1.70)	-0.008 (0.54)	-0.084 (2.37)
Certificate IV	0.008 (0.69)		-0.003 (0.37)	-0.027 (2.15)	-0.031 (1.45)
Certificate III	0.028 (2.17)	0.002 (0.02)		-0.033 (2.95)	-0.056 (2.86)
Certificate II	0.025 (0.81)	-0.035 (1.96)	-0.017 (1.48)		-0.034 (1.97)
Certificate I	-.031 (0.56)	-0.075 (2.51)	-0.038 (2.12)	-0.041 (2.39)	

In summary, the relative effects of undertaking higher-level VET qualifications is to increase the probability of being in full-time employment and to decrease the probability of being unemployed, while there is no effect on the probability of being in part-time employment.

Further study

Further study can be an ambivalent indicator, indicative of either success or failure. If doing a qualification increases the probability of further studies, it may be judged as a success, in that it has opened up opportunities for studies at a higher level than were previously available. On the other hand, if the further study is a second-best option for those who do not obtain employment, it may be seen as a failure. Neither of these two explanations for this outcome should be considered in isolation from other indicators of outcomes and previous research.

If further study is always undertaken at a higher level, one would expect that the reasons or opportunities for further study become reduced as the level of qualification increases. After all, having obtained the highest qualification, the opportunities for study at a higher level are exhausted. However, further study is not always at a higher level and even the highest vocational qualification does not preclude further study for a university degree. It is also well known that the lower-level certificates are often undertaken for the purpose of higher-level studies. There is a pattern to the results that is consistent with this simplistic view, in that most estimated effects are negative, meaning that a higher qualification reduces the probability of further study (tables 12 and 13). But the results where diploma and certificate IV are compared with certificate III depart from this pattern. Doing a higher-level qualification relative to certificate III has the effect of increasing the probability of further study. At the same time the negative effects of doing a certificate III qualification relative to certificates I and II are particularly large. This is suggestive of certificate III rather than higher qualifications being a terminal qualification. It is the highest level for the development of mainly practical skills that are directly applicable at work. Having reached this level there is little reason and few opportunities for higher-level qualifications. This has also been the predominant character of certificates III in the past, when most certificate IIIs were obtained as part of an apprenticeship in the traditional trades. On the other hand, the higher vocational qualifications are less ‘terminal’. Through the development of conceptual skills, these qualifications are just as suited to further study as they are for work, and even an advanced diploma is not the end of the road.

Table 12 Proportion in further study by qualification: all graduates

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.337	0.343	0.276	0.365	0.418

Table 13 Estimates of the effect of qualification i vs qualification j on the probability of being in further study for individuals who have competed either qualification i or j: all graduates

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		-0.035 (1.77)	0.037 (2.08)	-0.045 (2.08)	-0.021 (0.46)
Certificate IV	-0.001		0.062 (3.86)	0.001 (0.08)	-0.012 (0.42)
Certificate III	0.091	0.118		-0.086 (5.82)	-0.097 (4.09)
Certificate II	0.011	-0.026	-0.054		-0.040 (1.89)
Certificate I	-0.098	-0.062	-0.067	-0.017	

Benefits and relevance of training

The last two outcome indicators are qualitative in nature: the benefits and relevance of training. As regards job-related benefits, the estimates imply a sharp distinction between the three higher-level qualifications and the two lower. The probability of reporting a job-related benefit from doing a qualification higher than a certificate I is estimated to be 8 to 23 percentage points higher (table 15, column 5). The same results are obtained when the relevance of the training to the job is used as the outcome indicator (tables 16 and 17).

Table 14 Proportion reporting job-related benefits of training: graduates in full- or part-time employment after training

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.732	0.703	0.765	0.647	0.574

Table 15 Estimates of the effect of qualification i vs qualification j on the probability of reporting job-related benefits from training for individuals who have competed either qualification i or j: graduates in full- or part-time employment after training

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		0.002 (0.11)	-0.21 (1.16)	0.091 (3.87)	0.227 (4.10)
Certificate IV	0.050		-0.016 (0.99)	0.082 (4.00)	0.154 (4.44)
Certificate III	-0.033	-0.067		0.129 (8.01)	0.193 (6.92)
Certificate II	0.067	0.034	0.096		0.081 (3.25)
Certificate I	0.137	0.065	0.163	0.030	

Table 16 Proportion reporting training being relevant to job after training by qualification: graduates in full- and part-time employment after training

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
0.783	0.807	0.829	0.661	0.560

Table 17 Estimates of the effect of qualification i vs qualification j on the probability of the training being relevant to job after training for individuals who have competed either qualification i or j: graduates in full-time employment after training

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		-0.032 (1,78)	0.013 (0.80)	0.102 (4.53)	0.215 (3.99)
Certificate IV	0.051		0.021 (1.44)	0.109 (5.65)	0.201 (5.84)
Certificate III	-0.038	-0.055		0.148 (9.47)	0.214 (7.46)
Certificate II	0.081	0.011	0.109		0.122 (4.76)
Certificate I	0.168	0.074	0.214	0.108	

Module completers

The Student Outcomes Survey distinguishes between two main groups of completing students: graduates and module completers. About two-thirds of completers are graduates, while the remaining one-third are module completers. Since module completion is an integral part of the Australian VET system, there is a sense in which the issue we have analysed — the relative effect of the training at different levels — is just as relevant to module completers as it is to graduates.

Module completion also gives rise to another question that has received more attention among VET researchers: the effect of completion vis-à-vis partial completion, that is, whether graduates do better than module completers. The basis for this question is whether the high level of non-completion should be a case for concern or whether partial completion results in a proportionate improvement in outcomes.

In this section we examine whether, using matching methods, the relative effects of completing a module at different levels and the effects of completion versus partial completion can be credibly estimated.

The reason for the cautious formulation of the questions is that the identification of causal effects is conditional on the validity of the conditional independence assumption. In the case of graduates we argued that the assumption was plausible. The observed confounding variables were reasonable predictors of which qualification students choose, and matching on the propensity score led to a significant reduction in the differences between the treated and controls.

On a priori grounds there is no obvious reason why this should not also be the case for module completers. Module completers choose and are admitted to different qualifications on the basis of the same factors as graduates. Who is what — a module completer or a graduate — is determined after the event. Thus, there is no obvious problem with estimating the relative effect of completing modules at different levels. In the case of graduates versus module completers, or completion versus partial completion, however, the confounding factors we observe do not include any variables that are likely to affect the selection into treatment; that is, variables that determine who completes or does not complete. In fact, and as suggested by previous research, it is not clear that there are any observable pre-training factors that are good predictors of completion status: the reasons for module completers' non-completion are usually traced to what happens during training.[2]

To examine these issues we first look at the relative effect of module completion per se. In the case of graduates we sought to estimate the relative effects of qualifications. Since module completers do not complete a qualification, the effect has to be rephrased to the relative effect of completing modules at different levels. For illustrative purposes we restrict attention to one particular case: what is the effect on weekly earnings of completing modules at diploma level relative to modules at certificate III level?

The estimates of the bivariate probit from which the propensity score is derived are given in table A3 in the appendix. Compared with the corresponding results from graduates, the effect of the confounding variables is smaller and the predictive power of the model is correspondingly smaller. At the very least, these results raise questions about whether all the factors that influence both the selection into treatment (choosing diploma rather than certificate III) are included. Matching on the estimated propensity score does reduce the largest difference between the treated and controls, but overall the matching is quite poor. The median standardised bias (across all variables) is only reduced from 7.5 to 5.8% (table 18).

Table 18 Summary of the distribution of the standardised absolute bias of covariates before and after matching: diploma module completers vs certificate III module completers

	Percentiles	
	Before matching	After matching
10%	1.22	0.00
25%	3.46	1.98
50%	7.50	5.80
75%	10.90	11.63
90%	26.83	12.28

Proceeding to the estimates of the treatment effects we note that the average treatment effects are much smaller than for graduates (table 20). This largely reflects the smaller, before matching,

[2] Module completers is not an ideal term, making it difficult to avoid convoluted language. The main problem is that in one sense graduates are the completers (of a qualification), while module completers are the non-completers or partial completers.

differences in means in table 19. Indeed, the estimates of the treatment effects are quite close to the mean differences between the groups, indicating that the matched controls are not very different from the unmatched. The estimates are also imprecise. The certificate IV module completers is the only group who earn significantly more than they would have done had they undertaken modules at a lower level. In all other cases the relative effects are insignificant. The lack of significance means that the hypothesis, that the level at which module completers study has no effect on their earnings, cannot be rejected. This implication is nevertheless plausible. If completing one or more modules does not add much to the skills and knowledge a person already has, it is not surprising there are no differences according to the level of modules.

Table 19 Average weekly earnings by qualification studied for: module completers in full-time employment ($)

Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
967	1016	866	901	859

Table 20 Estimates of the effect of completing modules at level i vs level j on weekly earnings for individuals who completed modules at levels i or j: module completers in full-time employment after training

j i	Diploma and above	Certificate IV	Certificate III	Certificate II	Certificate I
Diploma and above		-17 (0.56)	60 (1.91)	-15 (0.38)	79 (0.86)
Certificate IV	17		81 (2.91)	42 (1.07)	216 (2.28)
Certificate III	167	109		-36 (1.02)	-80 (1.13)
Certificate II	175	118	-18		99 (1.73)
Certificate I	53	173	-50	-35	

Our analysis of the second issue — whether there is an effect of completing a qualification — is based on a comparison of graduates with module completers at the same level. More specifically we examine weekly earnings after graduating with a certificate III vs completing only one or more modules at certificate III level but here we provide the main findings. Estimating the propensity score from a bivariate probit model we find that only a few of the included variables have a large or significant impact. Thus level of schooling and highest qualification before training, which are the main determinants of which qualification a person undertakes, have little or no influence on completion status. This is reflected in a very poor predictive power of the probit model. It is also reflected in the distributions of the propensity score among the completers and module completers. As illustrated in figure 3, the distributions are almost identical, meaning that, a priori, module completers are just as likely to be graduates as the graduates themselves. The consequence of this is that matching on the propensity score does little to reduce the differences between the treated and controls. Before matching, the median standardised bias across all confounding variables was 5.3%. Matching on the propensity score reduced this figure to 2.6% (table 21).

Figure 3 Distributions of propensity scores

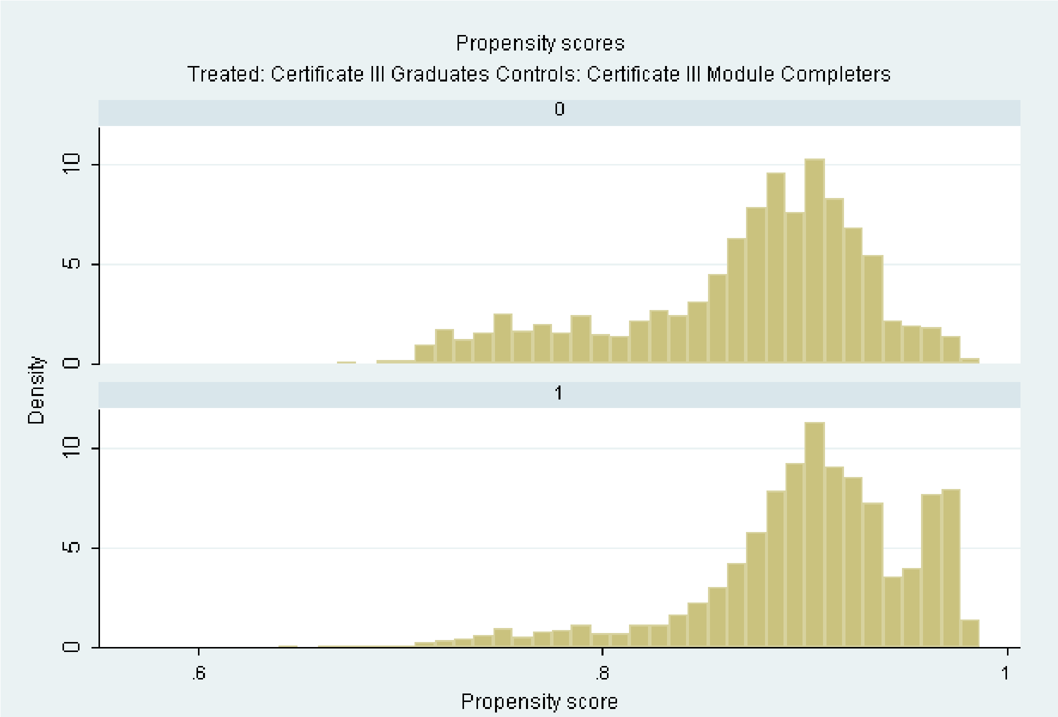


Table 21 Summary of the distribution of the standardised absolute bias of covariates before and after matching: certificate III graduates and module completers

	Percentiles	
	Before matching	After matching
10%	1.24	0.20
25%	2.47	1.14
50%	5.25	2.60
75%	8.80	3.23
90%	26.7	5.09

For the sake of completeness one set of matching estimates are given in table 22. The table reports the matching estimates of the effect on weekly earnings of graduating with a qualification (the treatment) versus completing one or more modules at the same level. According to the estimates, the average certificate III graduate increases their earnings by $68 per week by completing their qualification. As it happens, this is the only estimate that is significantly different from zero. This does not necessarily mean that there is no effect of completion but, in view of the poor quality of the matching, credible estimates of the effect requires richer data than available in the Student Outcomes Survey. The small estimates are nevertheless consistent with previous evidence. Dumbrell, de Monfort and Finnegan (2001) conclude that there is little difference in the outcome between graduates and module completers in diploma courses.

Table 22 Matching estimates of the effects of completing a qualification (graduates) vs completing one or more modules of the same qualification (module completers): individuals in full-time employment

		Mean weekly earnings		
	Treated: graduates Controls: module completers	Before matching	After matching	Average treatment effect on the treated (ATT)
Diploma and above	Treated	1026	1026	
	Controls	966	975	50 (1.90)
Certificate IV	Treated	1044	1044	
	Controls	1005	1046	-2 (-0.16)
Certificate III	Treated	863	863	
	Controls	865	795	68 (3.07)
Certificate II	Treated	817	817	
	Controls	904	817	0 (0.00)
Certificate I	Treated	747	747	
	Controls	864	842	-95 (-1.45)

The conclusion we draw from this analysis is that the relative effects of module completion and the effects of graduation versus module completion cannot be credibly estimated using the technique and data used in this paper. In the first case, the principal reason is that matching on the propensity score does not result in a well-balanced sample. In the second case the problem is that the critical conditional independence assumption is unlikely to be satisfied.

Conclusions

The aim of this paper was to illustrate the use of matching methods in VET research. To that effect the paper estimates the relative effect of VET qualifications on a range of outcomes using the data from the Student Outcomes Survey.

The results indicate that, for graduates, higher-level qualifications lead to better outcomes than do lower levels. Thus, higher-level qualifications increase earnings, improve employment outcomes and are more beneficial and relevant to jobs than lower-level qualifications. Higher qualifications also tend to promote further studies. While it is not possible to clearly distinguish between the relative effects of all five levels of VET, in most cases there is a three-level grouping of qualifications: high (certificate IV and above), medium (certificate III) and low (certificate II and below). Within these groups the relative effects of doing a higher qualification are small or imprecisely estimated. Between the groups, however, there are much larger differences and the estimates are more precise.

As regards module completers, we find that the technique and data used cannot be used to estimate either the relative effects of module completions at different levels or the effects of completing a qualification as opposed to only completing some modules. Credible estimates of these effects depend on more and better data and techniques with the ability to take account of unobservable effects.

The results for graduates broadly correspond to what can be directly observed from raw data and what has been found in previous studies using the data from this survey. In contrast to methods used previously, however, matching takes into account that outcomes are influenced by confounding factors and that these outcomes might depend on the choice of qualifications. To identify the effect of qualifications in these settings, the critical maintained assumption is that, conditional on the confounding factors, the outcomes are independent of the choice of qualifications. For that reason matching is often described as being a data-rich method: it requires that all variables that influence both the treatment and outcomes are observable and included in the analysis. By the standards of many other studies, the Student Outcomes Survey is not a rich dataset. In particular, it does not include socioeconomic characteristics, which are likely to influence the choice of education and training. It does, however, include many of the variables that influence both the selection of the level of qualification and the outcomes, notably, previous education and training. This is borne out by the large effect these variables have on the estimated propensity score and justifies matching the treated with controls using this score. Further justification is provided by the quality of the matching process. While the investigation of matching quality is not exhaustive, a number of indicators suggest that after matching we are indeed comparing like with like, and hence that the estimates represent the causal effects of qualifications.

In this study the focus has been on two parameters: the average effect on the treated (ATT) and the average effect on the untreated (ANT). For policy purposes it is generally held that the most relevant parameter is the former. This parameter gives information relating to the gain from an existing policy (here the VET system), given the choices people make. The latter can be seen as providing complementary evidence. This parameter tells us what the average effect of undertaking a higher-level qualification would be for those who undertook a lower-level qualification. This is of some relevance to a VET system that has been undergoing a shift from lower- to higher-level qualifications for some time, a shift that is likely to continue into the future.

The finding of an average positive effect from completing a higher-level qualification does not imply that everybody benefits. It is quite possible, even likely, that those who already have a high-level qualification do not gain much benefit from another qualification, at any level. Previous empirical evidence using a regression model suggests that this is indeed the case (Karmel & Nguyen 2007). In principle, matching techniques like regression analysis can be used to estimate the respective average effects for groups defined by their previous education and training. However, this option has not been pursued in this study. As the sample is split in smaller and smaller groups, the estimates become less precise and eventually little additional insight is obtained. This problem is clearly evident in Karmel and Nguyen (2007), who found that higher qualifications for those already well qualified yield a lower return than lower-level qualifications. This result is difficult to rationalise and should not be used as a reason for restricting VET.

References

Ball, K & Lamb, S 2001, *Participation and achievement in VET of non-completers of school*, Research report 20, ACER, Melbourne.

Bryson, A, Dorsett, R & Purdon, S 2002, 'The use of propensity score matching in the evaluation of labour market policies', Working paper no.4, Department of Work and Pensions, London.

Caliendo, M & Hujer, R 2006, 'The microeconometric estimation of treatment effects — an overview', *Allgemeines Statistisches Archiv*, Bd. 90, pp.197—212.

Caliendo, M, Hujer, R & Thomsen, S 2007, *The employment effects of job creation schemes in Germany*, Discussion paper no.512, IZA, Bonn.

Caliendo, M & Kopeinig, S 2008, 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys,* vol.22, no.1, pp.31—72.

Dehijia RH & Wahba, S 2002, 'Propensity score matching methods for nonexperimental causal studies', *Review of Economics and Statistics,* vol.84, no.1, pp.51—61.

Dumbrell, T, de Monfort, R & Finnegan, W 2001, *Outcomes for VET diploma and associate diploma students*, NCVER, Adelaide.

Frolich, M 2002, 'Programme evaluation with multiple treatments', Discussion paper 2002—17, Department of Economics, University of St Gallen, Switzerland.

Fullerton, S 2001, *VET in schools: participation and pathways*, LSAY research report 21, ACER, Melbourne.

Imbens, G 2004, 'Nonparametric estimation of average treatment effects under exogeneity: a review', *Review of Economics and Statistics*, vol.86, no.1, pp.4—29.

Imbens, G & Wooldridge, JM 2008, *Recent developments in the econometric of program evaluation*, Discussion paper 3640, IZA, Bonn.

Karmel, T, Mlotkowski, P & Awodeyi, T 2007, *Is VET vocational? The relevance of training to the occupations of vocational education and training graduates*, NCVER, Adelaide,

Karmel, T & Nguyen, N 2007, *The value of completing a vocational education and training qualification*, NCVER, Adelaide.

Lechner, M 1999, 'Earnings and employment effects of continuous off-the-job training in East Germany after unification', *Journal of Business Economic Statistics,* vol.17, no.1, pp.74—90.

——2001, 'Identification and estimation of causal effects of multiple treatments under the conditional independence assumption', in *Econometric Evaluation of Labour Market Policies*, eds M Lechner & F Pfeiffer, Physica, Heidelberg, pp.1—18.

——2002, 'Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods', *Journal of the Royal Statistical Society*, series A, vol.165, no.1, pp.59—82.

Leuven, E & Sianesi, B 2003, 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing and covariate imbalance testing', viewed February 2012, <http://ideas.repec.org/cboc/bocode/s432001.html>.

Long, M, Carpenter, P & Hayden, M 1999, *Participation in education and training 1980—1994*, LSAY research report 13, ACER, Melbourne.

Rosenbaum, P 1987, 'The role of a second control group in an observational study', *Statistical Science*, vol.2, no.3, pp.292—316.

——2002, *Observational studies*, Springer Verlag, New York.

Rosenbaum, P & Rubin, D 1983, 'The central role of the propensity score in observational studies for causal effects', *Biometrika,* vol.70, no.1, pp.41—55.

Rubin, D 1974, 'Estimating causal effects of treatments in randomised and nonrandomised studies', *Journal of Educational Psychology,* vol.66, no.5, pp.699—701.

Sianesi, B 2004, 'An evaluation of the Swedish system of active labour market programs in the 1990s', *The Review of Economics and Statistics,* vol.86, no.1, pp.133—55.

Smith, J & Todd, P 2005, 'Does matching overcome LaLonde's critique of nonexperimental estimators?', *Journal of Econometrics,* vol.125, no.1—2, pp.305—53.

Stanwick, J 2005, *Australian Qualifications Framework lower-level qualifications: pathways to where for young people?*, NCVER, Adelaide.

——2006, *Outcomes from higher-level vocational education and training qualifications*, NCVER, Adelaide.

Stromback, T 2010, 'Earnings, schooling and vocational education and training', *Australian Journal of Labour Economics*, vol.13, no.3, pp.241—63.

Teese, R & Walstab, A 2008, *Social area differences in vocational education and training participation*, NCVER, Adelaide.

Appendix: Detailed results not reported in the body of the paper

Table A1 Bivariate probit estmates of the probability of treatment: graduates

		Diploma vs certificate III		Diploma vs certificate I	
Variable		Coefficient	t-value	Coefficient	t-value
	Male	-0.402	-14.17	-0.540	-8.57
Area	Regional	-0.278	-9.26	-0.294	-4.33
	Remote	-0.617	-9.12	-0.967	-7.99
Age	20–24	0.354	5.04	1.670	11.76
	25–29	0.471	5.63	1.711	9.94
	30–34	0.576	6.59	1.755	9.66
	35–39	0.752	8.64	1.974	10.77
	40–44	0.774	8.89	1.564	9.12
	45–49	0.687	7.87	1.592	9.03
	50–54	0.663	7.36	1.660	9.13
	55–59	0.536	5.31	1.480	7.46
	60–64	0.358	2.67	1.365	4.95
	65 +	0.008	0.03	0.335	0.79
Start of training relative to leaving school	Within 12 months	0.195	2.26	0.859	5.31
	More than 12 months	0.140	1.65	0.248	1.47
	Prior relevant skill or experience	0.095	3.03	0.400	6.06
Prior labour force status	Unemployed	-0.230	-2.91	-0.199	-1.19
	Not in the labour force	0.113	1.54	-0.116	-0.8
	Indigenous	-0.137	-1.31	-0.765	-4.56
	English not spoken at home	-0.024	-0.52	0.043	0.4
Spoken English	Well	-0.283	-3.42	-0.242	-1.32
	Not well	-0.322	-1.54	-0.970	-3.07
Level of schooling	Year 11	-0.229	-5.08	-0.345	-3.62
	Year 10	-0.405	-10.13	-0.345	-3.95
	Year 9 or below	-0.561	-6.54	-0.603	-3.62
Prior qualifications	Bachelor degree	0.631	12.23	0.226	2.12
	Advanced diploma	0.928	11.18	0.712	3.76
	Diploma	0.915	17.33	0.592	5.21
	Certificate IV	1.232	23.93	0.874	7.65
	Certificate III	0.431	10.05	0.541	5.28
	Certificate II	-0.193	-3.39	0.158	1.11
	Certificate I	-0.293	-2.36	-0.811	-3.8
	Other certificate	0.052	0.65	0.215	1.18
	Certificate of competency	0.096	0.8	0.004	0.02
	Statement of attainment	-0.059	-0.43	-0.451	-1.71
	Pre-vocational	-0.078	-0.62	-0.163	-0.6
	Other vocational	-0.107	-0.71	0.355	0.88
	Constant	-1.275	-12.93	-0.710	-4.26
Number of observations		11 962.000		3 475	
LR chi2		2 451.220		1 112.8	
Prob >chi2		0.000		0	
Pseudo R2		0.186		0.346	

Table A2 Balance of the treated and matched controls before and after matching

Variable		Sample	Diploma vs certificate III				Diploma vs certificate I			
			Means		Standard-ised bias	Reduction in bias	Means		Standard-ised bias	Reduction in bias
			Treated	*Controls*			*Treated*	*Controls*		
	Male	Unmatched	0.459	0.632	-35.3		0.459	0.711	-53	
		Matched	0.459	0.461	-0.4	98.8	0.459	0.422	7.7	85.5
Area	Regional	Unmatched	0.332	0.446	-23.5		0.332	0.427	-19.7	
		Matched	0.332	0.345	-2.6	89	0.332	0.324	1.7	91.6
	Remote	Unmatched	0.033	0.073	-17.9		0.033	0.134	-36.8	
		Matched	0.033	0.030	1.6	91.3	0.033	0.047	-4.9	86.8
Age	20-24	Unmatched	0.208	0.342	-30.2		0.208	0.101	30.1	
		Matched	0.208	0.207	0.4	98.7	0.208	0.177	8.7	71.2
	25-29	Unmatched	0.136	0.112	7.1		0.136	0.068	22.6	
		Matched	0.136	0.141	-1.7	76.1	0.136	0.148	-4.3	81
	30-34	Unmatched	0.109	0.078	10.8		0.109	0.059	18.1	
		Matched	0.109	0.109	0.1	98.9	0.109	0.106	1.4	92.3
	35-39	Unmatched	0.127	0.074	17.7		0.127	0.050	27.6	
		Matched	0.127	0.123	1.4	92.1	0.127	0.111	5.7	79.4
	40-44	Unmatched	0.126	0.074	17.3		0.126	0.111	4.8	
		Matched	0.126	0.123	0.9	94.6	0.126	0.153	-8.3	-71.9
	45-49	Unmatched	0.116	0.078	12.9		0.116	0.081	11.8	
		Matched	0.116	0.124	-2.8	78	0.116	0.131	-5	57.4
	50-54	Unmatched	0.091	0.064	10.1		0.091	0.066	9.3	
		Matched	0.091	0.086	2	80.5	0.091	0.099	-2.9	69.3
	55-59	Unmatched	0.043	0.040	1.4		0.043	0.050	-3.2	
		Matched	0.043	0.045	-0.9	38.7	0.043	0.048	-2.3	26.4
	60-64	Unmatched	0.014	0.018	-3		0.014	0.015	-0.8	
		Matched	0.014	0.013	1.1	63	0.014	0.007	6.1	-704.9
	65 +	Unmatched	0.002	0.004	-4.6		0.002	0.010	-10.7	
		Matched	0.002	0.002	0	100	0.002	0.001	0.9	91.5
Start of training relative to leaving school	Within 12 months	Unmatched	0.111	0.244	-35.2		0.111	0.178	-19.1	
		Matched	0.111	0.102	2.3	93.4	0.111	0.064	13.4	29.8
	More than 12 months	Unmatched	0.867	0.695	42.4		0.867	0.583	67.1	
		Matched	0.867	0.882	-3.7	91.3	0.867	0.916	-11.7	82.6
	Prior relevant skill or experience	Unmatched	0.738	0.655	17.9		0.738	0.536	42.8	
		Matched	0.738	0.732	1.1	93.6	0.738	0.742	-1	97.6
Prior labour force status	Unemployed	Unmatched	0.023	0.055	-16.5		0.023	0.068	-21.6	
		Matched	0.023	0.019	2.2	86.9	0.023	0.030	-3.2	85.2
	Not in the labour force	Unmatched	0.033	0.052	-9.3		0.033	0.132	-36.5	
		Matched	0.033	0.024	4.7	49.6	0.033	0.021	4.6	87.3
	Indigenous	Unmatched	0.015	0.025	-6.9		0.015	0.069	-27.3	

Variable		Sample	Diploma vs certificate III				Diploma vs certificate I			
			Means		Standard-ised bias	Reduction in bias	Means		Standard-ised bias	Reduction in bias
			Treated	*Controls*			*Treated*	*Controls*		
		Matched	0.015	0.011	3	56.1	0.015	0.015	0.2	99.4
	English not spoken at home	Unmatched	0.163	0.122	11.9		0.163	0.122	11.7	
		Matched	0.163	0.152	3.3	72.3	0.163	0.152	3.2	72.8
Spoken English	Well	Unmatched	0.040	0.033	3.8		0.040	0.033	3.9	
		Matched	0.040	0.037	1.7	56.6	0.040	0.018	12	-205
	Not well	Unmatched	0.004	0.004	-0.7		0.004	0.015	-11.5	
		Matched	0.004	0.003	1.1	-53.2	0.004	0.004	0	100
Level of schooling	Year 11	Unmatched	0.112	0.146	-10		0.112	0.191	-22.2	
		Matched	0.112	0.102	3.1	68.8	0.112	0.092	5.7	74.5
	Year 10	Unmatched	0.150	0.260	-27.6		0.150	0.271	-29.9	
		Matched	0.150	0.136	3.4	87.7	0.150	0.125	6.1	79.5
	Year 9 or below	Unmatched	0.020	0.048	-15.2		0.020	0.061	-20.8	
		Matched	0.020	0.014	3.3	78.4	0.020	0.020	0	100
Prior qualifications	Bachelor degree	Unmatched	0.166	0.071	29.9		0.166	0.097	20.5	
		Matched	0.166	0.174	-2.4	92	0.166	0.167	-0.1	99.5
	Advanced diploma	Unmatched	0.048	0.015	19.1		0.048	0.013	20.5	
		Matched	0.048	0.032	9.6	49.9	0.048	0.053	-2.8	86.2
	Diploma	Unmatched	0.153	0.049	35.1		0.153	0.063	29.5	
		Matched	0.153	0.154	-0.2	99.3	0.153	0.177	-7.6	74.2
	Certificate IV	Unmatched	0.191	0.043	47.3		0.191	0.053	43.2	
		Matched	0.191	0.193	-0.6	98.8	0.191	0.129	19.3	55.4
	Certificate III	Unmatched	0.167	0.158	2.5		0.167	0.094	21.8	
		Matched	0.167	0.164	0.9	66.1	0.167	0.215	-14.4	34
	Certificate II	Unmatched	0.043	0.133	-32.2		0.043	0.059	-7.5	
		Matched	0.043	0.047	-1.5	95.4	0.043	0.043	0	100
	Certificate I	Unmatched	0.006	0.030	-17.7		0.006	0.084	-38.1	
		Matched	0.006	0.006	0.5	97	0.006	0.005	0.5	98.7
	Other certificate	Unmatched	0.025	0.041	-9		0.025	0.026	-1	
		Matched	0.025	0.026	-0.6	93.4	0.025	0.025	-0.4	57.9
	Certificate of competency	Unmatched	0.009	0.017	-6.6		0.009	0.020	-8.7	
		Matched	0.009	0.009	0	100	0.009	0.011	-1.5	83.2
	Statement of attainment	Unmatched	0.007	0.014	-6.9		0.007	0.023	-13.3	
		Matched	0.007	0.006	0.7	90	0.007	0.010	-2.6	80.6
	Pre-vocational	Unmatched	0.007	0.023	-12.8		0.007	0.017	-8.5	
		Matched	0.007	0.008	-0.3	97.8	0.007	0.021	-12.9	-51.8
	Other vocational	Unmatched	0.006	0.013	-8		0.006	0.008	-3.2	
		Matched	0.006	0.008	-2.5	68.4	0.006	0.002	3.8	-17.3

Table A3 Bivariate probit estmates of the probability of treatment: module completers

Variable		Diploma vs certificate III		Diploma vs certificate I	
		Coefficient	t-value	Coefficient	t-value
	Male	-0.0551	-0.75	-0.066	-0.41
Area	Regional	-0.4337	-5.59	-0.553	-3.2
	Remote	-0.9582	-5.61	-1.430	-4.41
Age	20–24	0.2171	1.25	0.914	2.19
	25–29	0.2942	1.43	0.541	1.14
	30–34	0.0917	0.43	0.857	1.64
	35–39	0.4162	1.97	0.892	1.83
	40–44	0.3490	1.65	0.569	1.23
	45–49	0.2746	1.27	0.290	0.62
	50–54	0.0911	0.4	0.089	0.18
	55–59	0.2307	0.92	-0.223	-0.43
	60–64	0.1555	0.5	-0.593	-1.05
	65 +	0.3398	0.62		
Start of training relative to leaving school	Within 12 months	0.6782	2.65	1.482	3.38
	More than 12 months	0.5156	2.08	1.132	2.58
	Prior relevant skill or experience	-0.0812	-1.04	0.142	0.85
Prior labour force status	Unemployed	0.0225	0.12	-0.048	-0.13
	Not in the labour force	0.4513	2.34	0.690	1.53
	Indigenous	0.1062	0.41	-0.544	-1.32
	English not spoken at home	0.2339	1.9	0.687	1.98
Spoken English	Well	-0.4499	-2.18	-1.151	-2.5
	Not well	-0.4805	-1.05	-2.088	-3.13
Level of schooling	Year 11	-0.2977	-2.61	-0.293	-1.24
	Year 10	-0.5736	-5.33	-0.254	-1.05
	Year 9 or below	-0.4455	-1.87	-0.705	-1.63
Prior qualifications	Bachelor degree	0.2070	1.68	0.601	2.36
	Advanced diploma	0.4348	1.75	1.033	1.55
	Diploma	0.7381	5.25	0.676	2.48
	Certificate IV	1.0198	7.21	1.599	4.15
	Certificate III	0.5516	4.59	0.836	3.24
	Certificate II	0.0764	0.49	0.755	1.82
	Certificate I	-0.3053	-0.9	-0.464	-0.81
	Other certificate	0.0910	0.44	0.386	0.93
	Certificate of competency	-0.1404	-0.38		
	Statement of attainment	-0.0075	-0.02	-0.704	-1.14
	Pre-vocational	-0.5222	-0.95	-0.622	-0.63
	Other vocational	0.1509	0.48	0.343	0.64
	Constant	-1.0940	-4	-0.649	-1.39
Number of observations		1500		569	
LR chi2 (37)		221.87		164.75	
Prob > chi2		0		0	
Pseudo R2		0.1179		0.3251	

NVETR Program funding

This work has been produced by NCVER under the National Vocational Education and Training Research and Evaluation (NVETRE) Program, which is coordinated and managed by NCVER on behalf of the Australian Government and state and territory governments. Funding is provided through the Department of Industry, Innovation, Science, Research and Tertiary Education.

The NVETRE program is based on priorities approved by ministers with responsibility for vocational education and training. This research aims to improve policy and practice in the VET sector. For further information about the program go to the NCVER website <http://www.ncver.edu.au>. The author/project team was funded to undertake this research via a grant under the NVETRE program. These grants are awarded to organisations through a competitive process, in which NCVER does not participate.

National Centre for Vocational Education Research Ltd
Level 11, 33 King William Street, Adelaide, South Australia
PO Box 8288, Station Arcade, SA 5000 Australia
Telephone +61 8 8230 8400 Facsimile +61 8 8212 3436
Website www.ncver.edu.au Email ncver@ncver.edu.au