1 Introduction

1.1 Background

The beginnings of the Australian vocational education system can be traced back to the early 19th century when the Mechanics Institutes were established in Hobart (in 1827) and Sydney (in 1833). Within a few years, Adelaide, Newcastle and Brisbane had established similar institutions. Vocational education became more formalised when, prompted by the need for skilled labour, New South Wales (1828 and 1844) and Tasmania (1844) passed a series of apprenticeship laws (Goozee, 2001). Mechanics institutes were also established in Victoria in the 1830s and 1840s, followed by schools of mines in Ballarat and Bendigo in 1870 and 1871, respectively (Anderson 1998). Queensland established its first mechanics institute in North Brisbane in 1849. Population growth led to rapid expansion of the need for a skilled workforce, and beginning from 1890 the New South Wales government began to increasingly assume responsibility for technical education while Victoria operated a more decentralised system.

As Australia developed, the need for a skilled workforce became ever more apparent. While technical education suffered financially in the first half of the twentieth century, mainly due to the Great Depression and two world wars, the post-World War Two period saw rapid population growth and a corresponding need for technical education. By the 1960s, several inquiries had determined that the technical education system was undervalued and under-financed, and recommendations from these inquiries included the increase of expenditure on technical education as well as the renaming of trade schools as technical institutions or colleges. In 1973, following the pivotal Kangan report (1974), the newly named TAFE system became a distinct entity in the national education sector. Despite the original convention of

state-administered vocational education, from the 1970s onward, resource constraints within the state governments' budgets and the increasing financial burden of running the system effectively led to growing Commonwealth involvement (Goozee, 2001). Further reform ensued in the mid-eighties, with the adoption of a uniform system of award nomenclature across the states. Since the 1990s, a multitude of reforms have affected the TAFE sector, including the introduction of competency-based training and assessment, a national system of qualifications, and the opening of the previously monopolistic market to private training providers (Rumsey, 2002). In the twenty-first century, schools, universities, industry and private training providers all form part of the competitive national VET market. Most recently, many TAFE colleges have abandoned the use of the 'TAFE' acronym from their institution titles in an effort to better position themselves in the competitive national and international training market (Martin et al., 2012).

Today, the publicly-funded vocational education and training sector represents expenditures in excess of $7.5 billion. Of this sum, state governments fund $3.35 billon, $2.06 billion is funded by the Federal Government, and the remainder is funded by fee for service, student fees and other ancillary trading income (NCVER Financial Information, 2010). Naturally, all these stakeholders have an inherent interest in producing desirable outcomes in return for financial and administrative input. These outcomes can be loosely grouped into two distinguishable categories of performance indicators: efficiency and effectiveness. The quantification of efficiency is important in centrally-planned or funded systems as knowledge about efficiency can be used to inform decisions about allocating funds between institutes. Effectiveness, on the other hand, encompasses a range of potential performance measures which enable the assessment of whether the students and stakeholders receive value from their investment in vocational education. Effectiveness measures also provide transparency in the provision of the educational service.

Performance measurement has become topical in non-profit and government organisations in recent years. The reasons for this are manifold, but include the assessment of the impact of improvement initiatives, national and international quality awards, changing external demands, increasing competition, and the power of information technology (Neely 1999). In post-compulsory education, performance measurement can be used by governments as a method of managing higher education via selective intervention based on performance indicators. This intervention is often combined with the allocation of funds and can therefore be viewed as threatening by the institutions (Sizer et al., 1992). However, it is reasonable for stakeholders to expect that there are some measures that allow the assessment of educational outcomes in relation to financial and administrative input as well as permit inter-institutional comparisons.

The aim of this research portfolio is to develop a framework that enables comprehensive performance measurement in the Australian TAFE[1] sector. To this extent, efficiency and effectiveness will be defined, determination of their constituent elements will be provided, a methodology for quantifying these elements will be developed, and benchmarking between individual providers will be facilitated. Special emphasis will be placed on the controversial issue of completion rates and a specific measure of the benefit of completion over non-completion for various groups of students will be derived.

The expected benefits of this portfolio are threefold. Firstly, this research aspires to make a genuine contribution to academic knowledge in the field of performance measurement in vocational education. Secondly this portfolio aims to introduce and apply econometric methods in vocational education research that have hitherto not been used in this area. Finally, and in line with the aims of a professional doctorate in education, the research presented in

[1] In the context of this portfolio, TAFE institute refers to TAFE institutes, TAFE divisions of a university, skills institutes and polytechnics.

this portfolio intends to be highly relevant to contemporary issues in vocational education administration. As such it is hoped that some of the research results presented here may lead to concrete and actionable policy recommendations and implementations.

1.2 The research problem

The modern vocational education and training (VET) sector is a fast-transforming environment. Rapid changes in the nature of work, technology, labour markets, and the impact of globalisation require constant policy and institutional responsiveness, which in turn requires up-to-date methods for measuring the effects of those new policies. The resulting changes in the Australian VET system have elicited increasing amounts of VET-related research. Despite this, it has been claimed that insufficient research exists which analyses problems that are complex organisationally and socio-economically. This may be due to inadequate research resources and/or insufficient research timeframes, as well as reluctance to venture into sensitive policy areas (Chappell, 2003) when, for instance, institutional funding may be impacted by specific research outcomes. It has also been claimed that it is difficult to obtain direct measures of impact of VET research (Stanwick et al., 2009), particularly within a limited timeframe, due to the complexities of the VET system which appears to be driven more by ideological and/or economic considerations than by attention to research information (Harris & Clayton, 2010). The research presented in this portfolio is intended to at least partially address these issues.

VET research in Australia can be categorised in two ways: *Institutional research* aimed at making policy recommendations and assessing the impact of new policy initiatives as well as internal research conducted by an institution to inform its own development, strategies and decision making, and *academic research* driven primarily by the quest for knowledge and development of

research methodologies. The primary purpose of this research portfolio is to contribute to some aspects of both of these research approaches, namely to incorporate and adapt existing quantitative methodologies for a suite of performance assessment measures, as well as to make some concrete actionable policy recommendations.

Firstly, it is hoped that this research helps to inform policy and practice. Effective evidence-based public administration requires data and tools to utilise the data in order to assess successes and failures of policy. This study will assess several aspects of the current state of public VET provision and defensibly quantify the performance of individual service providers relative to others in the Australian context.

Secondly, there is a need for the development of new quantitative methods in vocational education research. Cameron (2010) examined the papers published in the *International Journal of Training Research* (the Journal of the Australian Vocational Education, Training and Research Association (AVETRA)) from 2003 to 2008 in respect to the methodologies employed and could only classify 8% of papers as using quantitative research. Furthermore, presentations at the annual AVETRA conferences in 2007 and 2008 included only 4% and 8% quantitative talks respectively. Given this analysis, it appears that quantitative methods in Australian VET research are somewhat underrepresented. This represents a rather unfortunate reality as TAFEs, private VET providers, and various regulatory and policy research institutions do collect copious amounts of administrative and survey data which are often accessible from the public domain without being analysed and evaluated. Recent decades have seen an increasing trend toward the collection and analysis of numerical data. While there are some distinct benefits to be gained from qualitative research, the availability of quantitative data enables us in this research portfolio to address the underrepresentation of quantitative research in the VET sector. Specifically,

the research aims to draw a broad picture of various aspects of institutional performance in the Australian TAFE system (ATS) of vocational education.

In order to better delineate performance into coherent sub-categories we will make a distinction between effectiveness and efficiency and deal with them separately. Furthermore, we will subdivide the topic of effectiveness into two classes: on the one hand we will study a number of performance measures that relate to the evaluation of effectiveness at the institutional level. On the other hand we will consider a single effectiveness measure, completions in VET, at the system level. This portfolio will thus be comprised of three substantive papers : Paper 1 will investigate how institutional effectiveness can be determined using data from various survey and administrative collections. This paper will be named 'A framework of objective and subjective performance indicators'. Paper 2, titled 'On the efficiency of Australian TAFE institutes' will encompass a study of technical efficiency at the institutional level. Finally, Paper 3 will contain the effectiveness study at the systemic level, entitled 'Completion in VET — How beneficial is it, and who completes?'

A framework of objective and subjective performance indicators (Portfolio paper 1)

Whereas efficiency can often be defined in a straightforward manner as some variant of an output to input ratio, the notion of effectiveness is usually more difficult to capture. In the context of this portfolio, effectiveness will refer to a broad set of indicators that evaluate quantitatively a number of outcomes of individual institutions in the ATS. These indicators range from subjective student-based measures (typically considered to be 'soft' measures[2]), such as student satisfaction, willingness to recommend their institution, sense of achievement etc. to objectively measurable outcomes. This study endeavours to

[2] Such 'soft' measures operate without a rigorous quantitative foundation. For a detailed discussion of soft measures, see Forrester (1994)

ascertain summary measures of individual outcomes, to statistically validate these summary measures and segment them by student groups of interest (for example, graduates, module completers etc.) and to adjust them for the influence of institutional characteristics, including demographics of the student body, location, multi campus status, qualification level (certificate 1—4, diploma), and study mode (part time/full time) and socioeconomic background of its student body.

Objectively measurable indicators of effectiveness consist of two themes: The first theme could be broadly termed as student destinations. As such this theme will deal with those measures which relate to post-study effects, such as labour market outcomes (for example, time to find a job, labour market participation rate, salaries, occupational status of job attained etc.), and the uptake of further study.

In this paper we will use administrative and survey data from a specific reporting period, and will give emphasis to the development of quantitative methodologies that can be used by VET practitioners in an applied setting. Specifically, we are interested in answering a number of research questions.

The Student Outcome Survey contains a number of subjective outcome measurements, ranging from 19 individual satisfaction questions dealing with specific aspects of students' training, over sense of achievement and willingness to recommend the institution to others.

1) What type of summary measures can be created from this volume of data to create a meaningful composite index that enables inter-institutional comparison? How can the efficacy of these summary measures be validated?

Published data from the Student Outcome Survey are based on response rates below 40% but much lower still for some individual institutions/student groups. It seems reasonable to hypothesize that the large non-response introduces

significant selection bias into the survey, skewing most of the reported data. The comparative analysis of the effectiveness of individual institutions is further complicated by demographic, educational and administrative differences between institutions. We thus ask:

2) How can this bias be quantified and addressed in subsequent analyses? How can summary measures obtained from Q1 be utilized to give an accurate reflection of student's satisfaction with their training, while taking into account demographic composition and circumstances specific to individual institutions? How do TAFE institutes compare based on the measures developed?

Some more tangible measures of the effectiveness of TAFE institutes relate to labour market outcomes and patterns of further study. To complete our inquiry into TAFE effectiveness, we aim to examine:

3) What are the institutional effectiveness outcomes after demographic and institutional characteristics are taken into account? Do subjective effectiveness measures from Q1 relate to these objective outcomes, and can we use the subjective and objective measures developed in this paper to create a composite measure of effectiveness? Can we establish performance profiles against our effectiveness indicators?

On the efficiency of Australian TAFE institutes (Portfolio paper 2)

'Efficiency' is generally considered a quantitative measure of the relationship between inputs and outputs for a given process. In this study the input is more specifically defined as the resources that are employed by VET institutions to produce educational outcomes and post-educational (for example, labour market outcomes, salaries etc.) outcomes. Such resources may be monetary in nature, such as government funding for ongoing expenditure, funding for capital costs, as well as students' fees. In addition to financial inputs, input resources can be supplemented by data with respect to the characteristics of

individual TAFE institutes, such as their number of campuses, degree of remoteness, and profile of their faculty and support staff (including qualifications, part-time/full-time status and ratio/number of support staff).

Government and the public have an inherent interest in the efficiency of TAFE institutes, as it is good policy to produce satisfactory educational outcomes with scarce public funding. In this study, we employ parametric and non-parametric methods to quantify the efficiency of Australian TAFE institutes. Results from both quantitative approaches will be compared and analysed with respect to their similarities and differences. We look into the statistically significant drivers of efficiency and determine whether (and if so, which) specific policy recommendations could be derived from such analyses. The study also analyses if differences in efficiency can be differentiated from disparities in efficiencies attributable to the varying sizes of individual TAFE institutes. In this portfolio paper, our aim is to answer the following research questions:

4) Efficiency can be parameterized in various ways, depending on which outputs and inputs are considered. Which inputs and outputs provide a meaningful way to characterize efficiency in Australian TAFE institutes?
5) Given that there are different methodologies available to quantify institutional efficiency and that each of these methods has distinct advantages and disadvantages: What technique (for example, parametric or non-parametric) is more suitable to determine institutional efficiency, given the limited number of potential input and output variables? To what extent do parametric and non-parametric methods correlate?
6) As we can expect disparities in efficiency between institutions it is of interest what the drivers of efficiency are and if these drivers are uniform between different types of efficiency. We want to explore: How is institutional efficiency distributed within the Australian TAFE system, and

what variables drive this efficiency? Can a typology be developed that characterises efficient institutions?

Completion in VET – How beneficial is it, and who completes? (Portfolio paper 3)

Our second effectiveness portfolio paper theme considers completions at the system level. This theme is more comprehensive and thus deserving of a portfolio paper to itself. Completion rates have recently been the subject of an interesting debate. Published research by the National Centre for Vocational Education Research (NCVER) suggests that effective completion rates are below the completion rates claimed by some industry practitioners (Ross, 2011). There is some ambiguity that is inherent in published completion rates which stems from the fact that these rates are estimated via sophisticated models (in this specific case via Markov Chains, see Mark & Karmel 2010) rather than directly calculated. This is due to the way the data are submitted from service providers to NCVER. The study proposed here will seek to bring more clarity to this issue and propose an alternative way of estimating completions.

There can be little doubt that completion rates are a key indicator of TAFE performance. However, what remains unresolved is whether there actually is a measurable benefit in completing. In our third portfolio paper, we develop a methodology that addresses this problem.

Finally, one curious result of the TAFE Student Intentions Survey run in 2011 was that an overwhelming majority of students intended to complete the qualification they were enrolled in. This pattern contrasts with the aforementioned comparatively low completion rates. We attempt to create a link between responses from the Student Outcomes Survey to responses from the Student Intentions Survey to determine the actual outcomes of their

education. A model is then constructed to determine if there are significant factors that can be used to predict completions.

Recent research has indicated comparatively low completion rates in the vocational education sector, ranging from 13 to 48%, depending on field of study (Mark & Karmel, 2010). This contrasts with results from the Student Intention Survey, which indicated a generally high commitment (93%) of commencing students to complete the qualification they are enrolled in (NCVER, 2010). Completion rates are an obvious measure of performance. However, the issue of completion rates has recently been the subject of some controversy in the VET research community (Ross, 2011). Measuring completions is problematic, as the format in which institutional data is currently being reported has so far precluded a straightforward calculation of completion rates. We will develop a more standardised measure that enables inter-institutional comparison of completion rates. The specific answers we want to find in this third portfolio relate to these issues:

Completion rates in VET are notoriously low. It would be reasonable to speculate that low completion rates may result from some students dropping out of programs after they have obtained the skills they were interested in.

7) Is there is a quantifiable benefit to completing a qualification? If so, how does completion benefit several predefined segments of the student population (for example, qualification level, age group, socioeconomic background, field of education, prior education etc.)?

In addition, we are interested the practical relevance of the findings from the previous questions and will ask in our final research question:

8) What are the characteristics of the students who fail to complete their qualifications? How can we predict (and thus target) them?

All the research questions posed at the outset of this study deal with the quantification of performance in the Australian TAFE system. The main research theme of this study can thus be summarised as follows:

> TAFE institutes occupy an important role in Australian post-secondary education. Their stakeholders include federal and state governments, employers and industry skills councils, and students enrolled in programs of vocational education. Each of these stakeholder groups has a legitimate interest in the performance of TAFE institutes, but those interests may diverge between different groups. This study seeks to provide patterns and relationships that could help to meet the needs and expectations of multiple stakeholder groups and define the components to enable a comprehensive assessment of TAFE performance. This will be accomplished via the development of a suite of methodologies that evaluate a range of aspects of TAFE efficiency and effectiveness.

As such, we are aiming to define and analyse a number of measures that can be loosely categorised into the three clusters of effectiveness, efficiency, and completions. Each of these three clusters will provide an umbrella for the three individual portfolio papers.

The unit of analysis for most of the above research questions will be the institution. However, those questions dealing with intention and completion, as well as the benefit of completion will mainly focus on the individual student or groups of students (students of a certain qualification level, age group, employment status etc.). This doctoral research is presented in portfolio form, and therefore will comprise the above three substantive, self-contained journal-style articles. While the articles will be able to stand on their own, they will be integrated by a linking paper plus conclusion under the theme of 'Institutional efficiency and effectiveness in Australian TAFE institutes'. This linking paper (paper 4) will integrate the findings from the substantive papers 1 to 3, highlight research results, provide recommendations, and discuss potential implications for educational administration and policy.

Following from the research questions posed in the preceding paragraphs, this portfolio is presented under the following headings:

A Introduction to this portfolio

B Portfolio paper 1: 'A framework of objective and subjective institutional performance indicators'

C Portfolio paper 2: 'On the efficiency of Australian TAFE institutes'

D Portfolio paper 3: 'Completion in VET: Who completes, and how beneficial is it?'

E Linking paper: 'Efficiency and effectiveness of Australian TAFE institutes'

The substantive analysis from papers 1 to 3 will be structured as shown in the flow chart below, with green, blue, and orange shading representing the three portfolio themes of efficiency, objective and subjective effectiveness, and completions.

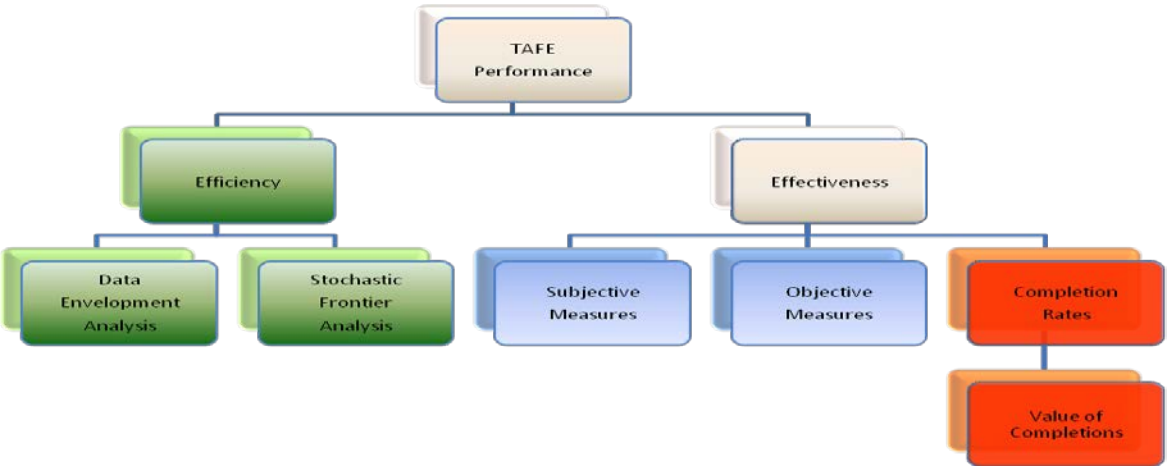

Figure 1.1 Structure of substantive focus of papers 1 to 3

1.3 Definition of terms and acronyms

In the context of this study, the following terms are defined as outlined below:

CHAID: Chi squared automatic interaction detection

Completion rates: is the ratio of those who complete a qualification over those who originally enrolled

DEA: Data Envelopment Analysis; a linear programming technique for efficiency analysis

Effectiveness: Set of outcome parameters which define the 'performance' of a TAFE. These parameters can be further divided into 'subjective effectiveness', encompassing student evaluations, and 'objective effectiveness', dealing with measurable education outcomes such as completion rates, labour market outcomes etc.

Efficiency: Efficiency refers to the relationship of output of educational and related parameters and input of tangible resources into public VET education providers, with the highest efficiency scoring unit assigned an efficiency of 1.

Graduate: A graduate is a respondent to the SOS who has completed a VET qualification (for example, certificate 1—4, diploma etc.). While the definition of graduate and module completer appears to be straightforward, there is some ambiguity in applying them (specifically in student self-identification), which has tangible consequences in their treatment in the SOS. This will be addressed in the thesis.

Module completer: A module completer is a respondent to the SOS who has completed at least one module of a VET qualification, but is believed to have left the VET system

SFA: Stochastic Frontier Analysis

SIS: Student Intentions Survey

SOS: Student Outcome Survey

VET: Vocational Education and Training

1.4 Prior research in the three portfolio areas

The notion of performance measurement in public institutions may appear to have become topical only in fairly recent times. Some basic forms of performance measurement however occurred as early as in the mid-eighteenth century. In 1754, the Pennsylvania hospital began collecting patient outcome data. While additional and isolated incidents of the measurement of institutional performance continued to occur throughout the eighteenth and nineteenth century, the first systematic attempt at performance assessment via a standardized measurement technique was implemented in 1910 in Massachusetts with the introduction of the tracking of hospital patients with the aim to determine whether their treatment was effective (McIntyre et al., 2001). From the early 1900s onward, performance measurement of public institutions became more commonplace, albeit often confined to selected areas. The healthcare sector played a central role in the development of performance measurement as evidenced by 87% of all performance frameworks implemented in this area (Klassen et al., 2010). The advent of the information age in the late 1970s featuring increasing capacities to collect and process data set off a new era in performance measurement. The development of performance indicators became more commonplace also in post-secondary education, albeit predominantly in the university sector and focussed chiefly on research outputs (Ramsden, 1991). Increasing financial pressures lead to an intensified focus on efficiency and effectiveness which in turn led to greater emphasis on the examination of measures of performance (Higgins, 1989). More recently, research engaged in creating and analysing performance indicators has used more advanced econometric and mathematical models in some educational sectors (Worthington, 2001). Performance measurement in Australian post-secondary education has been criticised as too centralised and

a low degree of political trust of central authorities in institutions to do their work effectively and efficiently (Woelert & Yates, 2014).

Research in performance measurement in the Australian TAFE system has been comparatively scant. Performance in the TAFE sector has chiefly been evaluated at the system level. Performance indicators devised and evaluated include students' participation and achievement in VET and training, student achievements, student outcomes, employer engagement and satisfaction with VET, and VET system efficiency (DEEWR, 2011a). It is only recently that an interest in performance measures at the institution level has emerged. Research into such indicators has mostly been driven by government policies and initiatives in order to give regulators tools for overseeing the sector, creating transparency for effective markets, and provide governments with information about the success of their programs (Karmel et al., 2013). Consequently, there is a dearth of academically motivated research in this area and it is one of the purposes of this portfolio to contribute to filling this void. Generally, existing TAFE performance indicators evolve around the notion of effectiveness (such as student satisfaction, labour market outcomes etc.). The three constituent papers of this portfolio are meant to illuminate three specific aspects of contemporary performance measurements. These three aspects include conventional effectiveness measures (Paper 1), measures of efficiency (Paper 2), and course completions (Paper 3). While the first two papers deal with measurement issues at the institutional level, Paper 3 will look at course completions at the individual and system level.

Paper 1: A framework of objective and subjective institutional performance indicators

Analysis of institutional effectiveness is usually the domain of policy-driven research and it is therefore no surprise that most of the previous published research has been performed by government advisory organisations. Such government-initiated research into the performance of the VET sector can be

traced back to 1974 when the Whitlam government established the Kangan Committee to provide advice to the Minister of Education. Since then successive organisations from the TAFE clearinghouse, TAFE National Centre for Research and Development (renamed NCVER in 1992), ANTA Research Advisory Council and The National Research and Evaluation Committee (NREC) have engaged in VET research and published studies of the performance of VET (Loveder & Guthrie, 2008). However, the development and usage of performance indicators grew only slowly. Guthrie (1991) reported that by 1990 Victoria, Tasmania and the Northern Territory had made only little or no progress in developing performance indicators.

Of the performance indicators that have been developed, the most commonly used types are those measuring student satisfaction. Data for satisfaction indicators are collected by institutions and centrally (for example, the SOS), and are reported in institutional reports. There has been some research into the validity of the satisfaction component in questionnaires. Ward (2008) employed a structural equation model to examine constituent components of student satisfaction. While this analysis was done using data from only one TAFE institute, it defined several underlying traits that constituted student satisfaction and underlined the need to collect satisfaction data in a uniform fashion from all TAFEs. Bontempo and Morgan (2001, 2003) evaluated the validity of the grouping of satisfaction questions in the SOS and found that the question items did indeed measure the traits they were meant to capture.

More in-depth analysis of satisfaction data, however, has been scant, specifically with respect to how TAFE institutes perform relative to each other. Curtis (2010) performed the only comprehensive comparative analysis of satisfaction with TAFE education. Curtis employed Rasch analysis (a variant of item response theory) with 2009 SOS data and used the resulting indicators to build linear models of the various satisfaction dimensions. These models

included various covariates (such as socioeconomic and institutional traits) in order to adjust institutional results for such variables. He then analysed TAFE institutes in respect to the differences between their predicted versus observed values and constructed a ranking that enabled inter-institutional comparisons.

It is clear that there is a shortage of reliable comparative analysis of performance indicators in vocational educational institutions. While the Australian federal government in recent years has advocated more transparency with respect the comparability of performance indicators (as evidenced by the MySchool and, more recently, MyUniversity websites[3]), in reality there is still a shortage of reliable information. It is thus the purpose of this paper to develop new methodologies for the creation of indicators, as well as creating additional performance measures.

Paper 2: On the efficiency of Australian TAFE institutes

While efficiency and productivity of private enterprises can often be determined via traditional accounting methods, efficiency of public institutions that produce non-financial outcomes has been more difficult to evaluate in these terms. Two prominent methods have been developed in the late 1970s to overcome this problem.

Charnes at al. (1978) developed the prototype of data envelopment analysis, a non-parametric method to estimate a production frontier. This method represents a non-parametric variant of linear programming which deterministically constructs an efficiency frontier over the data. This method has been utilised in numerous efficiency assessments in public administration situations, such as in hospitals (Nayar & Ozcan, 2008), prisons (Butler & Johnson 1997), military (Charnes et al., 1985), law enforcement (Carrington et al. 1997), tertiary education (Abbott & Doucouliagos, 2003), and other areas.

[3] <www.myschool.edu.au> and <www.myuniversity.gov.au> provide parents and students with access to information about Australia's primary, secondary, and tertiary education providers. These websites display information on student demographics and a range of other institutional statistics.

In Australia, this method has been applied in various tertiary education settings, predominantly universities (Abbott & Doucouliagos, 2003; Carrington et al., 2004), TAFE institutes in Victoria (Abbott & Doucouliagos, 1998, 2002) TAFE institutes in Queensland (Abbott & Doucouliagos, 2000), and TAFE institutes nationwide (Fieger, 2010; Fieger et al., 2010). Generally, the findings resulting from data envelopment analysis in the Australian higher education sector indicate that the sector is relatively efficient and that productivity growth was superior to most other sectors of the economy. However, the analyses often had some significant conceptual limitations, as they were either locally constrained, or relied on very limited input parameters. The analyses performed in the TAFE sector by Abbott and Doucouliagos (1998, 2000, 2002) omitted a number of institute-specific variables (for instance, a measure for the operational complexity of administrating a given institute) that may have significant impact on their cost structure and thus flow-on effects on their efficiency. Furthermore, these existing studies often used raw data as an input into their model, which does not take into account things like the demographic composition of the student body. The study proposed here intends to address these shortcomings.

Research aims and settings of stochastic frontier analysis (SFA) are somewhat similar to data envelopment analysis. Applications of SFA, however, are often geared toward commercial situations, such as the evaluation of entire industries, for example, banking (Kraft, Hofler & Payne, 2006) or the steel industry (Wu 1995). There is some limited literature available that applies SFA to educational settings, mostly related to universities. Examples are Izadi et al. (2001) and Stevens (2005) who both determined the suitability of the method to estimate efficiencies, but found it challenging to apply the correct parameterisation in order to calculate the appropriate cost function.

To the knowledge of this author, stochastic frontier analysis has not been used previously to estimate the efficiency of TAFE institutes in Australia. The proposed study will not only attempt to fill this gap, but also aim to validate the estimated efficiency scores resulting from the SFA via analytical comparison with the results from data envelopment analysis. Such comparisons have been conducted before, although outside the field of higher education. Wadud and White (2000) calculated agricultural efficiencies and found that varying input measures impacted differently on efficiencies depending on method used. Cullinane et al. (2006) investigated parametric and non-parametric efficiency scores for identical scenarios in the transport industry and found a high correlation of results of both methods. The proposed study will determine how these methods compare in the context of Australian vocational education.

Paper 3: Completion in VET: Who completes, and how beneficial is it?

There are two distinct types of completion rates in the VET sector, module completion rates and course completion rates. While the calculation of module completions is straightforward (they generally begin and finish within the same reporting period), the calculation of course completions is much more complex. This is partially due to the lack of a unique student identifier, which makes it difficult to link a specific enrolment to a given completion across more than one reporting period.

Research in this area has been scant and only one study has attempted to develop a solution for this problem (Mark & Karmel, 2010: Bednarz ,2012, conducted a follow-up study using the same methodology as Mark & Karmel), though there has been controversy as some industry practitioners believed the Mark and Karmel study substantially underestimated completion figures (Ross, 2011). One possible issue with the work done by Mark & Karmel is the assumption that the probability of attrition of individual students remains constant over the period of their enrolment. It seems reasonable to doubt that

this assumption universally holds and the intention is to create an alternative method of deriving course completion indicators.

Another aspect of the presented research is the apparent discrepancy between intended completions as indicated by the recent Student Intentions Survey, and actual completion rates. This represents a very new field of investigation, as student intentions to complete have not been canvassed on a national level in the VET sector prior to 2011. Existing research focuses on the impact of other extraneous factors that impact on the probability to complete. For instance, Karmel and Mlotkowski (2010) investigated the impact of wages on completions. They found that there was a relationship between wages and completion, but that this relationship was not consistent across the various combinations of qualification type and gender.

Finally, the proposed study will aim to investigate the value of completing a VET qualification. The aforementioned completion rates indicate that a significant proportion of students place only little value on completing their studies. Karmel and Nguyen (2007) have therefore attempted to quantify the value of completing a qualification. They examined the financial and labour market pay-off to actually completing a VET qualification and found that the largest completion benefit accrued to those who entered with lower qualifications (for example, certificate I and II). Those with higher entry qualifications on the other hand got no wage benefit from completing. While the Karmel and Nguyen research remains the sole study dealing with the pay-off to completion in the VET sector, a study researching the same issue in the university context has found improved labour market outcomes even for non-completers, while completion had the expected positive impact on salaries (Marks, 2007).

The portfolio study presented here will determine pay-off to completion for a whole range of outcome indicators and identify student groups for whom completion is particularly beneficial.

1.5 Research theory, methodology, and design

This research develops methodologies that facilitate the measurement of the performance of educational institutions and determine how different performance measures relate to each other. The study relies mostly on secondary data that were aggregated to the institutional level. The traits analysed in the course of this research are measurable, for example, completion rates, attrition rates, and efficiencies, to name a few. Furthermore, the research aims to uncover what the predictors of these performance measures are, for example, what causes differences in performance, and how performance can be impacted. To achieve this aim, we perform a number of hypothesis tests utilising a number of different statistical techniques. Specifically, this is a quantitative data-driven study, with the aim of generalisable and replicable results. These considerations about the intended research illustrate that a positivist set of guiding assumptions is employed to conduct the study with a hypothetico-deductive model that underpins it.

A positivist framework essentially transforms social issues and dynamics to constructs that can be operationally defined using quantitative measurement, which can then be mathematically/statistically related to each other. While there is a wide array of distinct epistemological definitions of positivism (Halfpenny, 1982), in contemporary language the notion of positivism usually relates to inquiry-based on empirical, scientific principles (Friedman, 1979). Unlike in the constructivist framework, positivists do not concern themselves with the possibility of multiple interpretations of social phenomena. While many critical constructivists in the social sciences deliberately introduce

ideology and an agenda into their research, positivists go to great lengths to eliminate forms of bias (Schrag, 1992).

In a qualitative study, one does not begin with a theory to test or verify. Instead, consistent with the inductive model of thinking, a theory may emerge during the data collection and analysis phase … or be used relatively late in the research process as a basis for comparison with other theories (Creswell, 1994, quoted in Anfara & Mertz, 2006). On the other hand, research guided by the positivist paradigm often starts out with a clear set of testable hypotheses and the structure and layout of the resulting paper can often be determined in the early stages of the research process. There are obvious advantages of the positivist approach to the research presented here. These include the availability and accessibility of quantitative data, and the replicability and generalisability of research results which is a specifically desired outcome of this study. However, the positivist framework has its own limitations. Especially in the social science realm, the positivist approach has been criticized as being overly mechanistic and reductionist with a view of nature which defines social life in measurable terms rather than in terms of human experience, and excludes the notions of choice, freedom, individuality, and moral responsibility (Cohen et al., 2007).

Furthermore, on a methodological level, correlational/survey research, as it is presented here, has inherent problems in respect to various biases (for example, sampling/questionnaire/selection), as well as issues which emerge from the dichotomy of correlation and causation. While there are some biases in the data employed in this study, we will apply methodologies that have the goal to minimise the practical effects of such biases. Correlation versus causation issues are dictated by the available data and the results of the research produced in this portfolio are inherently correlational. To make definite statements on the causality of the relationships between variables in

this research would require conducting controlled experiments which, due to the complexity and large number of required study subjects, is not feasible.

The restriction on causal inferences necessitates the application of a broad spectrum of statistical techniques at varying stages of the research process in an effort to compensate somewhat for those limitations. Statistical methods employed include data envelopment analysis, stochastic frontier analysis, cluster analysis, correlation, principal component analysis, OLS regression, logistic regression, Heckman selection models, discriminant analysis, and chi squared interaction detection (CHAID).

Paper 1: A framework of objective and subjective institutional performance indicators

Institutional effectiveness is evaluated mostly by using data from the Student Outcomes Survey (SOS). This portfolio paper utilises data from the 2011 SOS, which is of significance in two ways.

First, odd years are 'big' SOS years. This stems from a design feature of this survey, which requires that the accuracy considerations in odd years are based on institutional requirements, whereas in even years this requirement is for accuracy at the state level. As a result, odd years use an augmented sample of potential respondents, which roughly triples the sample size (from 100,000 to 300,000) in big years. Second, the other two portfolio papers will use data also relating to 2011. Having research results from all three portfolio papers relating to the same time period will produce more reliable results when discussing the outcomes of the three studies in the linking paper.

Particular attention is given to the preparation of data for this analysis. Two specific issues will be taken into consideration to address data quality and usability. These are the addressing of bias and the creation of composite satisfaction scores. While the SOS employs a semi-post-stratification measure in order to address some demographic bias, there is still some inherent

response related bias. This bias is somewhat difficult to quantify, but it appears reasonable to speculate that those students who were less content with their TAFE experience are less likely to respond. This study investigates how to address this bias, and specifically employs the two-step consistent estimator method developed by Heckman (1979).

Presently, the SOS contains a large number of questions in respect to student satisfaction with their training. These questions are very specific and are impractical to use. This portfolio paper employs principal component analysis to segment these questions into coherent categories and then evaluate methods to create summary measures that can be used easily in subsequent analysis.

The main analysis and creation of institutional indicators relies on standard quantitative social science research methods, such as OLS and logistic regression. Here we use specific outcome measures (for example, a certain type of satisfaction) as dependent variables, and enter co-variates (for example, the measures we want to control for such as demographics and institutional characteristics) as well as the effect coded institutions into our model(s) in order to obtain institutional estimates for the outcome measure. This analysis is replicated for a whole host of outcome measures (and adapted to logit models for categorical outcomes). Finally, we examine how these performance measures relate to each other and whether a composite effectiveness measure can be developed. The data necessary for analysis in this portfolio paper are sourced from the 2011 Student Outcomes Survey.

Paper 2: On the efficiency of Australian TAFE institutes

Non-parametric data envelopment analysis has been applied in several post-compulsory education settings in Australia. As such, it is an established method to determine the efficiency of public sector organisations such as TAFE institutes. The method that will be developed in this study will differ in some aspects from previous implementations. While earlier DEA analyses of TAFE

institutes utilised very limited data inputs that did not usually account for unique circumstances of individual institutions, in this study auxiliary measures are developed that account for such circumstances (for example, number of different courses taught at individual TAFE institutes and other factors that affect the cost structure of a TAFE). Furthermore, in addition to outcomes such as the number of graduates, we introduce output measures that include labour market outcomes and analyse and compare the derived efficiencies. We also include module completers in the analysis which thus far have been omitted, despite the fact that they are numerically the largest output of TAFE institutes.

The second methodology employed to assess efficiency is stochastic frontier analysis. This is a competing, and yet very different technique. As a parametric method, it is based on specific distributional assumptions. Input and output measures used are identical to the DEA analysis, in order to enable a side-by-side comparison of the two methods. Results are then correlated and analysed with respect to which parameters drive the efficiency to what extent.

Several data sources are used for this analysis. Firstly, funding data for TAFE institutes, including capital cost and salary/other ongoing cost are obtained from data requests and published institutional annual reports. Secondly, student data concerning labour market outcomes etc. are derived from the 2011 Student Outcome Survey. Thirdly, additional data are obtained from NCVER's 'Students and Courses' 2011 administrative collection.

Paper 3: Completion in VET: Who completes, and how beneficial is it?

There are two methodological components to this paper: the estimation and quantification of the benefit of completion and the analysis of the rift between the intention to complete and actual completion. This study defines 'completion' of a qualification by using the implicit completion definition inherent in the Student Outcome Survey, for example, we define a 'graduate' as a completer and 'module completer' as a non-completer.

As completion rates in vocational education have been found to be so extraordinarily low, it could be suspected that many individual students do not deem completion very rewarding. We want to test this proposition for various segments of the student population on various post-education outcomes. These outcomes range from salaries, attained occupational status, and uptake of further study to occupational status. The data for this analysis are derived from the 2011 SOS. An original model is built with an outcome (for example, salaries) as a dependent variable. A number of independent variables (field of education, age sex, qualification level, state, ARIA [region], employment status before training etc.) are then used in a completer model to predict individual salaries, not only for completers, but also for non-completers. Then, an identical model for non-completers is run and to predict salaries for non-completers and completers. As a result we have two predicted salary values for each individual student: one for the completer scenario and one for the non-completer scenario. The benefit of completion for the outcomes (in this case salary) is then defined as the ratio of the two values. Graphs are then produced to visualise the impact of the benefit of completion over the student population.

The final step is the creation of a classification tree. The method used here is called Chi squared automatic interaction detection (CHAID), as originally developed by Kass (1980). Using this method, we construct a tree where each (non-terminal) node identifies a split condition, to yield an optimal prediction (for continuous dependent or response variables) or classification (for categorical dependent or response variables). The categories are then merged until only two categories remain for each predictor. The algorithm then proceeds and selects among the predictors the one that produces the most significant split. This method is perfectly suited to identify and visualise student groups of interest (for example, by age, sex, qualification level etc.) and to what extent they benefit from completing their qualification.

Results from the 2011 Student Intentions Survey (SIS) indicated large differences between intended completion and actual completions. In order to better understand the rift between intended and actual completions we developed a model that integrates information from the Student Intentions Survey into the Student Outcome Survey. We use the 2011 editions of both surveys and initially model the intention to complete in the Student Intentions Survey using a number of predictor variables also present in the Student Outcome Survey. Once the model coefficients are established we make an out of sample prediction of the intention to complete for students in the Student Outcome Survey. After then predicting the probability of completion in the Student Outcome Survey, we have two continuous measures for every student available indicating intention to complete and actual completion. A number of subsequent regression and CHAID analyses then evaluate to relationship between intention to complete and actual completion as the determinants of both.

The data necessary for analysis in this chapter were sourced from the Student Intentions Survey and the Student Outcome Survey.

1.6 Ethical considerations attached to this portfolio research

All analyses and investigations in this portfolio utilise secondary data. As such, survey data is drawn from publicly available collections, such as the Student Outcomes Survey and the Student Intentions Survey. Both of these surveys are designed, collected, and maintained by the National Centre for Vocational Education Research (NCVER). As far as this author is aware, these surveys collected their data in accordance with the principles of modern research ethics guidelines, specifically the use of informed consent, voluntary participation, and assurance of confidentiality. Additional data was sourced from NCVER's 'Students and Courses' collection, an administrative database of student enrolment records.

NCVER provides public access for researchers to its collections, under the guiding principle that access needs to protect the confidentiality of the data (NCVER Data Access Policy). NCVER complies with the *Privacy Act 1988* and the National Privacy Principles (NPPs) which form part of that Act, in its management of personal information and sensitive information related to race, religion, political affiliation or to any health information it may collect. NCVER follows these national principles in its collection, use, maintenance, disclosure and storage of such information (NCVER privacy policy). NCVER indicates that:

> External users are only granted access to national vocational education and training data holdings (unit records stored at NCVER) where they have demonstrated to the satisfaction of the National Training Statistics Committee (delegated to NCVER) that they have a research interest in information at this level, and sign an undertaking agreeing to abide by the data protocol.
>
> (NCVER protocols for collecting and reporting VET statistical information)

Financial information for the efficiency section of this portfolio (paper 2) study is sourced from institutional annual reports. These documents are ordinarily available on individual institution's web pages for public perusal. Where annual reports are not available, financial data is collected via data request at individual institutes. No issues in respect to intellectual property rights have arisen from the preparation of this portfolio.

2 A framework of objective and subjective institutional performance indicators: Portfolio paper 1

2.1 Structure

2.1.1 Introduction

Performance measurement is an essential means for evaluating the success of public policies and programs. In the vocational education and training (VET) system, performance indicators can provide regulators, policymakers, other stakeholders, and the institutions themselves with a means to monitor and evaluate policies and outcomes. This also enables greater accountability by providing the evidence needed to give the full range of stakeholders the confidence that public funds are being appropriately spent and policy objectives

are being achieved. Moreover, these indicators can help to benchmark institutional performance and enable comparisons between education providers as well as facilitate the identification of strengths and weaknesses of individual institutes. The Australian Government acknowledges these aims and has in the past introduced several policy initiatives to pursue them.

Performance indicators have evolved in response to national developments in the VET sector. The emergence of modern performance indicators can be traced back to the development of the Australian Vocational Education and Training Management Information and Statistical Standard (AVETMISS) in 1991. This policy laid the groundwork to provide a nationally consistent standard for the collection and analysis of VET data (NCVER, 2002). In 1994 the Annual National Report of the Australian National Training Authority (ANTA) contained the four performance measures: enrolment, funding per annual hours of contact, training completions and module completions (ANTA ,1994). These performance indicators aimed to measure efficiency and effectiveness. In 1996 three new indicators, participation rate, graduate destinations and satisfaction, and employer satisfaction were added to the suite of key performance measures (ANTA, 1996). In 2002 the Australian Quality Training Framework was introduced with the goal of implementing nationally agreed quality assurance arrangements for the training and assessment services delivered by training organisations (NCVER, 2013). There have been several revisions to its guidelines with the 2007 version turning the focus on quality skills outcomes, outcomes-based auditing, and continuous improvement.

More recently the federal government has embarked on a drive to increase the effectiveness of the VET sector and part of this move is the 'Greater transparency of the VET sector' reform (Transparency Agenda, 2012). One specific goal of this policy is to make information on the different training providers and the quality of their performance more widely accessible. It is

the aim of this research to contribute to this goal and develop a methodology which provides for objective and easily interpretable performance indicators that can be used by regulators, policymakers and students alike in order to assess differences between institutions.

Developing and implementing performance indicators in the VET sector is not an easy task. Sauvagot (2007) argues that four questions need to be considered before developing an appropriate set of performance measures. Firstly, what aspects do we want to measure in terms of the goals and objectives of the sector? Secondly, what can we measure given the sources currently available and their credibility? Thirdly, what are we actually going to measure (the suggested final set of agreed indicators) and fourthly, what should we be able to measure in the future. The data sources that are ordinarily used to develop performance indicators can be classified as input (for example, indicators are usually concerned with agreements, declarations and commitments as well as human, financial and physical resources), process (for example, indicators show the extent to which the players have acted to meet their responsibilities and commitments), and output (for example, relating to measuring the results achieved in relation to the objectives set) (Horne, 2007).

In this study we are primarily concerned with the development of performance indicators that deal with outputs. We will be using data from the Student Outcome Survey (SOS) as well as selected labour market and census data. The SOS, being the most comprehensive survey of Australian VET graduates and module completers, provides a rich data source which is already being used to report a number of training outcomes. There are, however, a number of issues which impede a proper interpretation of the raw data reported by the SOS. These issues range from matters arising from the survey methodology for collecting the data (such as the potential existence of non-response and selection bias and low internal validity capacity for inferring cause and effect

relationships) to the difficulty of comparing institutions which differ vastly in respect to student profiles, offered training programs, strategic intent and organisational and regional culture. A further problem that encumbers a straightforward comparison of institutes is the fragmented nature of the multitude of satisfaction items in the SOS. With respect to employment outcomes, how can we adequately compare institutes if we have data about their graduates' employment status and salaries, but do not have any information in the survey about the employment and income situation in the area to which students move after they have completed their training (as it seems reasonable to assume that employment outcomes are not only a function of the quality of the training, but also of local employment conditions)? Finally, for a comprehensive analysis of performance indicators it seems necessary to evaluate how the different measures relate to each other.

It is thus our main objective to create and evaluate performance indicators that account for these issues and go beyond simple descriptive statistics. Specifically, we are aiming to address the following questions:

- How can non-response bias in the SOS be quantified and addressed in subsequent analyses?
- What type of student satisfaction summary measures can be created from the Student Outcomes Survey to construct a meaningful composite index that enables inter-institution comparison? How can the efficacy of these summary measures be validated?
- How can these summary measures be utilized to give an accurate reflection of students' satisfaction with their training, while taking into account demographic composition and circumstances specific to individual institutions? How do TAFE institutes compare based on the measures developed?

- How do the measures developed in the previous questions relate to each other? Are there underlying constructs under which the indicators could be categorized?
- Can other subjective outcome measures derived from the SOS (such as the perception of achievement and the willingness to recommend the institution of training) be developed as performance indicators?
- What is the impact of TAFE institutes on labour market outcomes such as employment after training and post training salaries after demographic and institutional characteristics are taken into account?

The approach for this study is to define and describe the data utilized to answer the research questions at hand. This includes the creation or derivation of auxiliary variables such composite satisfaction scores and inverse Mills ratios for our selection models. We will then develop eight hierarchical regression models which will be adjusted for demographic, environmental and institutional variables as well as selection bias to create eight specific performance indicators. Finally, we will compare the indicators created, examine possible relationships between indicators, discuss their utility and investigate whether there are associations between indicators and some student demographics.

2.1.2 Data and methods

In this study we will assess the effectiveness of TAFE institutes[1] mostly by using data from the Student Outcomes Survey (NCVER, 2011b). One design feature of this survey is an enhanced accuracy provision in odd years due to institutional requirements, whereas in even years this requirement is for accuracy at the state level. As a result, odd years use an augmented sample of potential respondents, which roughly triples the sample size (from 100,000 to 300,000) in big years. We therefore use the 2011 wave of the SOS, as this represents the

[1] In the context of this study, the term 'TAFE institute' refers to TAFE institutes, TAFE divisions of a university, skills institutes and polytechnics.

most recent enhanced SOS dataset. In this study we will pay particular attention to the preparation of data for the analysis. Two specific issues will be taken into consideration to address data quality and usability. These are the addressing of bias and the creation of composite satisfaction scores.

While the SOS employs a post-stratification measure designed to address some demographic bias (for instance, fewer males than females and fewer younger than older students tend to respond to the survey), there is still some remaining response bias (for instance, if response rates differ by qualification level) that is not currently addressed. This remaining bias is somewhat difficult to quantify, but it appears reasonable to speculate that those students who were less content with their TAFE experience are less likely to respond. This study will investigate measures to address this bias. While there are some well-known approaches in survey statistics that deal with this problem, such as raking[2] (see Deville et al., 1993) or propensity scoring[3] (see Leow et al., 2004), in our analysis we will employ the two-step consistent estimator method developed by Heckman (1979).

Presently, the SOS contains 19 questions with respect to student satisfaction with their training. These 19 questions are very specific and are impractical to use. This portfolio paper will employ principal component analysis to condense these questions into coherent themes or constructs and then evaluate methods to create summary measures that can be used easily in subsequent analyses.

The main analysis and creation of institutional indicators will rely on standard quantitative social science research methods, such as mixed effects OLS and logistic regression. Here we will use specific outcome measures (for example, a certain type of student satisfaction) as dependent variables, and enter covariates (for example, the measure we want to control for such as

[2] Raking employs iterative post-stratification to match marginal distributions to known population benchmarks.
[3] Propensity scoring essentially gives larger weights to those respondents who exhibit similar properties to non-respondents.

demographics and institutional characteristics) as fixed effects as well as the institution (and, for instance, the local unemployment rate in employment models) as random effects into our model(s) in order to obtain the institutional solution (for example, estimates) for the outcome measure. This analysis will be repeated for a whole host of outcome measures (and adapted to logit models for dichotomous outcomes). Finally, we will examine how these performance measures relate to each other and whether there are any demographic, institutional, or educational characteristics which can be used to predict performance. All statistical analyses reported in this paper were performed using SAS.

2.1.3 Data characteristics and preparation

The Student Outcomes Survey (SOS) 2011 was undertaken as a stratified, randomly selected sample with survey responses weighted to population benchmarks from the National VET Provider Collection. As the estimates from the SOS are based on information provided by a sample rather than a population, they are subject to sampling variability; that is, they may differ from the estimates that would have been produced if all graduates or module completers had been included and had responded to the survey (NCVER, 2011).

The 2011 wave of the SOS had an original population of around 1.3 million students (NCVER 2011e) from which a sample of around 356,000 students was drawn (table 2.1).

Table 2.1 Composition of the intended 2011 SOS sample

	Sample	Percentage
Graduates	123,299	35.0
Module completers	232,388	65.0
Total	**355,687**	**100.0**

The sample was selected as a systematic sample of the student population, stratified by 75 study locations (consisting of the TAFE institutes discussed in this study, as well as a number of aggregated private training providers), six

fields of education (aggregated from the 12 Australian Standard Classification of Education broad fields, three categories for sex (male/female/unknown), and five age ranges (15—19, 20—29, 30—44, 45+, and unknown).

While at the time of sampling there was an initial assumption that respondents have the status of 'graduate' (someone who had completed a qualification) or 'module completer' (someone who had completed a part of a qualification and is thought to have left the VET system), this did not always turn out to be true and could only finally be determined after responses had been received. The 'module completer' assumption was particularly prone to error, as many students re-enrolled or graduated after NCVER received the population data for sampling, and were hence misclassified as 'module completer'. These students were flagged as 'graduate' or 'continuing student' in the survey file and in the case of 'continuing student' their records were usually omitted from analyses of student outcomes. A further result of this re-classification of initial 'module completers' was that their high proportion (65%) of the total population (table 2.1) declined to 34% in the final received sample. The final respondent numbers for the TAFE portion of the SOS used for the analyses in this paper can be seen in table 2.2. A table containing sample and population data for all institutes covered in this analysis is included in appendix 1.

Table 2.2 Final TAFE respondent numbers to the 2011 SOS

	Respondents	Population	Percentage
Graduates	48,237	424,550	66.0
Module completers	15,514	218,968	34.0
Total	**63,751**	**643,518**	**100.0**

In this paper we included an analysis of selection bias in respect to several outcome variables using the two step method proposed by Heckman (1979). For this, it was necessary to estimate the probability of getting a response from individual respondents. To calculate response probabilities, we used the entire SOS sample of responding and non-responding students (including

students with private providers). We did this to strengthen the statistical power of our prediction of receiving a response and in the knowledge that there were well known patterns of non-response (for instance, males being less likely to respond than females (Smith 1983), rural people being less likely than urban people (de Leeuw & de Heer, 2002)) that could be assumed to be uniform across training provider types. All subsequent analyses however were based only on students who attended one of the TAFE institutes.

Labour force data for this study have been sourced from the Department of Education, Employment and Workplace Relations (DEEWR, 2011). For this present paper we utilized the unemployment rate by labour force region from the September 2011 quarter. As the SOS contained the home postcode of individual students, ABS concordances have been used to aggregate postcodes to statistical local areas and link those areas to labour force regions. There were a comparatively small number of responding students (<1000) whose home postcode represented a post office box. Technically speaking, such postcodes were not linked to ABS geographical postcodes. In an effort to retain student response data for our analysis we recoded post box postcodes to the geographical postcode in which the individual post office was located.

The most recent available census data were used to create an income proxy for all Australian geographical postcodes. As the census collected income data by having individuals pick one salary bracket out of a possible ten, we used the mid-point of each salary range and weighted it for the total number of respondents in that range and then calculated a mean for every postcode. This mean income per week and postcode was then multiplied by 52.14 to determine an average annual income. While there were limitations to this method (for instance, the resulting mean may have been biased if the distribution of incomes within ranges was skewed), it was believed that this technique provided an acceptable proxy for annual income. This was due to

the same methodology being applied across the board, meaning that resulting income values should still have maintained a similar ranking between individual postcodes. Furthermore, when these income data were used in the subsequent prospective hierarchical regression analysis as a random effect, it was the difference between incomes by postcode that facilitated the usefulness of this random effect, rather than the absolute values. As with labour market data, we recoded post box postcodes to the geographical postcode in which the individual post office was located.

2.1.4 Satisfaction question groupings

The Student Outcome Survey is an annual national survey of Vocational Education and Training (VET) students. The survey aims to gather information on students, including their employment situation, their reasons for undertaking the training, the relevance of their training to their employment, any further study aspirations, reasons for not undertaking further training, and satisfaction with their training experience. The survey is aimed at students who have completed a qualification (graduates), or who successfully completed part of a qualification and then leave the VET system (module completers).

The assessment of student satisfaction with their training consisted of 19 individual questions and one summary question (see figure 2.1a). The teaching and assessment questions were based on questions asked in the Higher Education Course Experience Survey and the generic skills and learning experience questions were based on questions developed by Western Australia as part of the VET Student Survey (Bontempo & Morgan, 2001). These questions occupied a significant portion of the questionnaire (20 out of 56 questions). To date the focus has been on reporting only the overall satisfaction item. Use of the individual satisfaction questions has been limited, mainly due to their specificity, narrow scope and the number of measures.

The individual satisfaction questions were grouped under three themes: Teaching, Assessment, and Generic skills and learning experiences. While there had been some initial statistical validation of these three groupings, no significant recent analysis had been undertaken, and no summary measure of the constituent questions had been devised.

In our analysis of the satisfaction questions we aimed to statistically validate the grouping of the questions in the context of current surveys and developed a summary measure for each of the three themes to make the data more usable for subsequent analyses. We used principal component analysis to identify the underlying dimensions of the 19 satisfaction items and grouped the questions accordingly. Observations with missing values and observations that had the 'Not applicable' field ticked were very rare and were omitted from computations. Cronbach's alpha scores were calculated to assess the internal consistency reliability of the resulting component measures.

We then used three different approaches to derive composite scores to represent the components created. These methods were Rasch analysis, weighted composite averages and straight unweighted averages. We then determined the extent to which the newly established composite scores differed and which ones would be used in our subsequent analyses.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable
Teaching						
1 My instructors had a thorough knowledge of the subject content	☐	☐	☐	☐	☐	☐
2 My instructors provided opportunities to ask questions	☐	☐	☐	☐	☐	☐
3 My instructors treated me with respect	☐	☐	☐	☐	☐	☐
4 My instructors understood my learning needs	☐	☐	☐	☐	☐	☐
5 My instructors communicated the subject content effectively	☐	☐	☐	☐	☐	☐
6 My instructors made the subject as interesting as possible	☐	☐	☐	☐	☐	☐
Assessment						
7 I knew how I was going to be assessed	☐	☐	☐	☐	☐	☐
8 The way I was assessed was a fair test of my skills	☐	☐	☐	☐	☐	☐
9 I was assessed at appropriate intervals	☐	☐	☐	☐	☐	☐
10 I received useful feedback on my assessment	☐	☐	☐	☐	☐	☐
11 The assessment was a good test of what I was taught	☐	☐	☐	☐	☐	☐
Generic skills and learning experiences						
12 My training developed my problem solving skills	☐	☐	☐	☐	☐	☐
13 My training helped me develop my ability to work as a team member	☐	☐	☐	☐	☐	☐
14 My training improved my skills in written communication	☐	☐	☐	☐	☐	☐
15 My training helped me to develop the ability to plan my own work	☐	☐	☐	☐	☐	☐
16 As a result of my training, I feel more confident about tackling unfamiliar problems	☐	☐	☐	☐	☐	☐
17 My training has made me more confident about my ability to learn	☐	☐	☐	☐	☐	☐
18 As a result of my training, I am more positive about achieving my goals	☐	☐	☐	☐	☐	☐
19 My training has helped me think about new opportunities in life	☐	☐	☐	☐	☐	☐
Overall satisfaction with the training						
How would you rate, on average, your satisfaction with the overall quality of the training?						
20 Overall, I was satisfied with the quality of this training	☐	☐	☐	☐	☐	☐

Figure 2.1 Student satisfaction items in the Student Outcomes Survey

Source: Student outcomes survey questionnaire (NCVER 2011).

2.1.5 Satisfaction themes

The bank of satisfaction questions in the Student Outcomes Survey was based on questions developed for use in the Western Australian State Student Survey. These questions were developed by the Western Australian Department of Education and Training, which also undertook an initial statistical validation (for more on the history of the satisfaction questions see Bontempo & Morgan, 2001 and Sevastos, 2001). Western Australia used the bank of questions in 2003, and a modified version became a constituent part of the current national Student Outcome Survey in 2004.

While there have been several evaluations of the categorization of the satisfaction questions into the three main themes, and these have provided a statistical basis for question groupings over the history of the survey (Morgan & Bontempo, 2003), there has been scant progress towards creating summary measures beyond the initial categorization into the three current themes.

Using principal component analysis[4], we identified the underlying component structure of the 19 satisfaction items and grouped the questions accordingly. The eigenvalues of the correlation matrix of the initial principal component analysis are shown in table 2.3.

Table 2.3 Eigenvalues[1] of the correlation matrix (abridged)

Item	Eigenvalue	Difference	Proportion	Cumulative
1	10.310	8.050	0.543	0.543
2	2.260	1.199	0.119	0.662
3	1.061	0.420	0.056	0.717
4	0.641	0.089	0.034	0.751
5	0.552	0.100	0.029	0.780
:	:	:	:	:
18	0.217	0.042	0.011	0.991
19	0.175	0.009	1.000	

Note: 1 Eigenvalues represent a characteristic root of a correlation matrix that can be used to determine how many factors should be retained. Ordinarily eigenvalues >1 are used as a cut-off for determining the number of factors. However, in borderline cases such as displayed in table 2.3a further analysis is warranted.

[4] Principal component analysis is a multivariate technique that is used to find lower dimensional projections of higher dimensional data to capture a large amount of the original variability with fewer variables. It differs from Factor Analysis in maximizing the total variance whereas Factor analysis maximises the shared variance

While there were various ways of assessing the number of components that should ideally be retained, we applied parallel analysis as suggested by Horn (1965), which uses a Monte Carlo-based simulation to compare the observed eigenvalues with those obtained from uncorrelated normally-distributed variables. The visual inspection of the resulting graph (figure 2.2) indicated that three components should be retained. These three extracted components accounted for about 72% of the variance in the 19 satisfaction items.

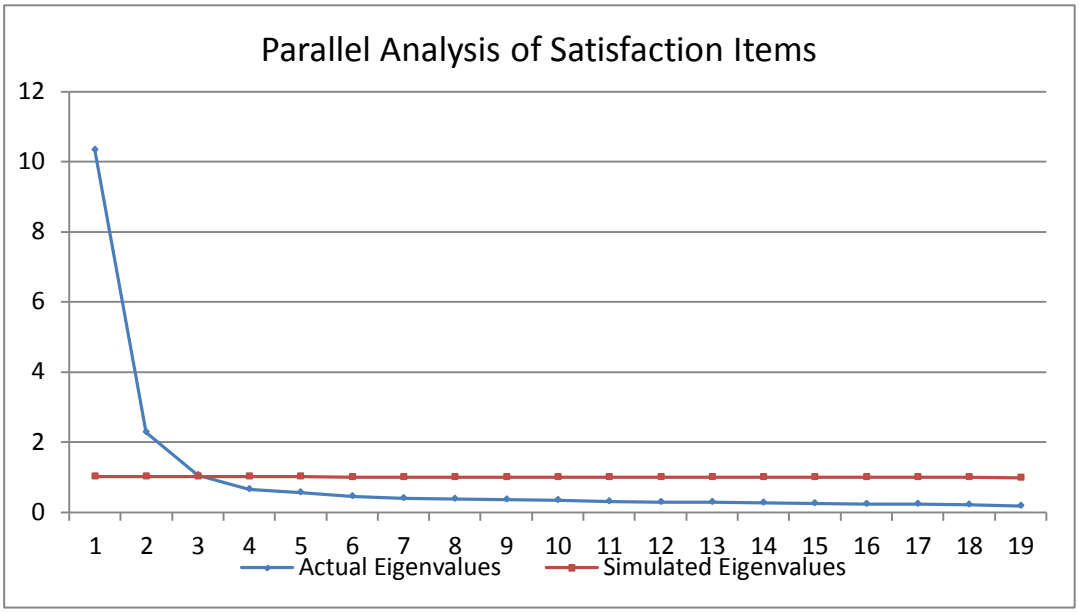

Figure 2.2 Eigenvalues based of parallel analysis

The pattern matrix resulting from the three retained components was then transformed via promax rotation. Non-orthogonal rotation was chosen as there are a number of studies which have shown that the components of satisfaction items are often highly correlated (see, for instance, Churchill et al., 1974 and Ragheb &Griffith, 1982) and therefore non-orthogonal rotation was methodologically more appropriate to deal with non-independent components. Indeed, our analysis confirmed the suspicion of highly correlated components (table 2.4)

Table 2.4 Correlation between factors

	Factor 1	Factor 2	Factor 3
Factor 1	1	0.52	0.56
Factor 2	0.52	1	0.64
Factor 3	0.56	0.64	1

It is very apparent that each single question unambiguously related to one particular component (shaded in table) and that the resulting three item clusters corresponded to the three thematic question groups from the survey (table 2.5). For example, those questions (1 to 6) that aligned with component 2 corresponded to the teaching block, those (7 to 11) aligning with component 3 corresponded to the assessment block, and those (12 to 19) aligning with component 1 corresponded to the generic skills and learning experience block of questions.

Table 2.5 Factor loadings after transformation using oblique promax rotation

	Factor 1 General	Factor 2 Teaching	Factor 3 Assessment
My instructors had a thorough knowledge of the subject content	-0.012	0.865	-0.032
My instructors provided opportunities to ask questions	-0.049	0.908	-0.023
My instructors treated me with respect	-0.043	0.887	-0.020
My instructors understood my learning needs	0.119	0.636	0.183
My instructors communicated the subject content effectively	0.059	0.726	0.155
My instructors made the subject as interesting as possible	0.133	0.656	0.128
I knew how I was going to be assessed	-0.061	-0.052	0.873
The way I was assessed was a fair test of my skills	0.025	0.064	0.826
I was assessed at appropriate intervals	-0.004	0.060	0.837
I received useful feedback on my assessment	0.105	0.122	0.669
The assessment was a good test of what I was taught	0.130	0.132	0.691
My training developed my problem solving skills	0.756	0.026	0.097
My training helped me develop my ability to work as a team member	0.794	0.021	0.032
My training improved my skills in written communication	0.835	-0.086	0.046
My training helped me to develop the ability to plan my own work	0.862	-0.051	0.037
As a result of my training, I feel more confident tackling unfamiliar problems	0.834	0.048	0.005
My training has made me more confident about my ability to learn	0.866	0.049	-0.026
As a result of my training, I am more positive about achieving my goals	0.863	0.059	-0.035
My training has helped me think about new opportunities in life	0.797	0.044	-0.032

Note: Shading indicates the question highly correlates with one particular factor.

We further evaluated the reliability of the three components using Cronbach's coefficient of internal consistency (table 2.6). All three components displayed excellent internal consistency as evidenced by a very high Cronbach's alpha statistic. None of the 'alpha if deleted' values exceeded the overall alpha score, which further reinforced the high reliability of the selected satisfaction components.

Based on the results of the principal component analysis and the review of the Cronbach's alpha scores, we concluded that the grouping of the satisfaction items into the themes (now components) of teaching, assessment, and generic skills and learning experiences in the Student Outcomes Survey was statistically justified.

Table 2.6 Descriptive statistics and coefficients of reliability

Question	N	Mean	Standard deviation	Alpha if deleted	Alpha
1	102,277	4.491	0.769	0.908	
2	102,277	4.514	0.750	0.904	
3	102,277	4.540	0.761	0.907	0.919
4	102,277	4.258	0.860	0.906	
5	102,277	4.296	0.866	0.899	
6	102,277	4.211	0.921	0.907	
7	100,609	4.231	0.829	0.901	
8	100,609	4.285	0.811	0.871	
9	100,609	4.259	0.810	0.874	0.901
10	100,609	4.131	0.966	0.887	
11	100,609	4.235	0.853	0.875	
12	94,659	3.919	0.917	0.936	
13	94,659	3.910	0.964	0.936	
14	94,659	3.714	1.027	0.937	
15	94,659	3.891	0.966	0.933	0.942
16	94,659	3.961	0.918	0.933	
17	94,659	4.031	0.925	0.932	
18	94,659	4.026	0.929	0.932	
19	94,659	4.069	0.943	0.938	

2.1.6 Comparison of composite measures

It seemed reasonable to speculate that the narrow scope of the individual satisfaction questions, along with the number of questions, had discouraged

their use in previous research. It was therefore desirable to have a composite score or summary measure for each of the three components that encapsulated the more general construct being reflected. This should be done by capturing the core information contained in the individual questions, while retaining as much information as possible. The result should be three component scores representing teaching, assessment, and general skills and learning experiences.

2.1.7 Rasch analysis

Rasch analysis is a one-parameter variant of item response theory and is used chiefly to analyse test scores or attitudes that are represented by Likert-type scales. The Rasch measurement model is employed to evaluate the fit of items to their intended scales and to generate individual scores and estimate the precision of those scores along an interval scale. The method also provides diagnostic information about the items and responses to them. Under item response theory, a set of items is assumed to reflect an underlying trait (such as satisfaction with teaching, assessment or learning) and responses to items are taken to indicate how strong individuals are on that trait and how easy or difficult it is to agree with an item reflecting that trait. A more thorough description of the method can be found in Bond and Fox (2007) and Fisher and Molenaar (1995).

We calculated the Rasch scores for this study with the help of the JMetrik 2.1 software, an open source computer program used primarily for psychometric analysis (JMetrik, 2013). This included analysing the constituent items from the three dimensions that were identified earlier in our principal component analysis. Observations with missing data were ignored. The reliability of the constructed Rasch item scales was exceptionally high (reliability scores of 0.9998 for teaching satisfaction, 0.9993 for assessment satisfaction, and 0.9996 for learning satisfaction). This was a consequence of the very large

number of observations and therefore resulting low standard errors of item estimates.

The weighted mean square errors for all three satisfaction dimensions were between 0.075 and 1.25, indicating compliance with the assumptions of the Rasch model (Wright and Masters, 1982). It should be noted that one of the main benefits of Rasch estimates is that the Rasch model creates a genuine interval scale, provided the assumptions are met.

2.1.8 Simple unweighted averages

As a second measure, we created a composite score for each of the three themes by calculating unweighted averages (often called 'unit-weighted scores', see Grice, 2001) for each individual. These mean scores were created even when individual responses to satisfaction questions were missing, for example, if the response to a question was missing the measure was calculated using the average of the remaining questions. This method thus maximised the use of the available data, while at the same time using the fewest administrative and computational resources.

2.1.9 Weighted averages

When using the above unweighted average scores, it can be argued that not all individual items contribute to the composite score to the same extent. It is useful to create a measure that accounts for the varying contributions of individual responses to the overall score. To create such a measure, we estimated component scores for the three identified dimensions. The scores had a mean of zero and a standard deviation of one, and represented the three themes of teaching, assessment and generic skills and learning experiences. Then we regressed the constituent satisfaction scores onto the component scores, with the aim of determining the strength of association of individual questions to the composite score. The resulting standardized regression coefficients provided a measure of the strength of the contribution

to the composite score. The composite scores for each component were calculated as:

$$Score_{weighted} = \sum_{i=1}^{I} Q_i W_i \quad (1)$$

where W are the respective weights derived by:

$$W_{qi} = \frac{B_i}{\sum_{i=1}^{n} B_i} \quad (2)$$

And Q_i represents the respective question for each component. The result represents the weighted average score for teaching satisfaction that has the same metric as the unweighted average score. One disadvantage of this method was that when a response for an individual satisfaction question was missing, a meaningful weighted composite score could not be calculated unless the missing response was imputed. Since response data for individual questions were only rarely missing (if satisfaction responses were missing they were usually missing for the entire respondent record), this issue was considered to be a negligible problem. Thus, where a missing item score occurred, the weighted average component score was missing.

2.1.10 Assessment of utility

As a result of the application of the above methodologies, we had available three different sets of composite scores for the three component themes. The basic descriptive statistics for the three summary measures can be found in table 2.7

Table 2.7 Descriptive statistics for composite scores scored used different methods

Variable	Scoring method	N	Mean	St. dev.	Min.	Max.
Teaching	Rasch	105,248	4.048	2.793	-5.223	7.214
	Unweighted mean	105,248	4.381	0.697	1	5
	Weighted mean	102,277	4.467	0.675	1	5
Assessment	Rasch	100,609	3.515	2.812	-5.587	7.085
	Unweighted mean	100,609	4.228	0.724	1	5
	Weighted mean	100,609	4.239	0.715	1	5
General teaching and learning	Rasch	94,659	2.805	2.892	-6.855	7.716
	Unweighted mean	94,659	3.940	0.800	1	5
	Weighted mean	94,659	3.932	0.808	1	5

While the unweighted mean and weighted mean scores appeared fairly similar, the mean and variation of Rasch scores were different. We therefore calculated correlations and Cronbach's alpha to examine the commonalities between the different methods and their reliability (tables 2.8, 2.9 and 2.10).

Table 2.8 Comparison between teaching composite scores

Calculation method	Rasch	Unweighted means	Weighted means
Rasch	1	0.953	0.912
		<.0001	<.0001
Means	0.953	1	0.973
	<.0001		<.0001
Weighted means	0.912	0.973	1
	<.0001	<.0001	
Alpha		Raw	0.719334
		Standardized	0.981310

Table 2.9 Comparison between assessment composite scores

Calculation method	Rasch	Unweighted means	Weighted means
Rasch	1	0.966	0.952
		<.0001	<.0001
Means	0.966	1	0.982
	<.0001		<.0001
Weighted means	0.952	0.982	1
	<.0001	<.0001	
Alpha		Raw	0.743067
		Standardized	0.988674

Table 2.10 Comparison between general skills and learning composite scores

Calculation method	Rasch	Unweighted means	Weighted means
Rasch	1	0.977	0.975
		<.0001	<.0001
Means	0.977	1	0.998
	<.0001		<.0001
Weighted means	0.975	0.998	1
	<.0001	<.0001	
Alpha		Raw	0.776537
		Standardized	0.994355

The main finding here was that correlations between the three methods were exceptionally high, with minimum correlations of 0.91 between Rasch scores and the weighted means method in the teaching and assessment themes (tables 2.8 and 2.9) and reaching practically 1 between unweighted means and weighted means method in the generic skills and learning experiences theme (table 2.10).

Cronbach's raw alpha scores encompassing the three aggregation methods were 0.72 for teaching, 0.74 for assessment, and 0.78 for generic skills and learning. The values suggested a very high degree of inter-item correlation[5]. Cronbach's standardised alpha scores could be interpreted as an indicator of inter-item covariance. In the three themes, the standardised values for teaching, assessment, and generic skills and learning experiences were all around 0.99. This suggested a very similar distribution of Rasch scores, means, and weighted means. Taken together, Cronbach's raw and standardised scores indicated strong internal consistency between Rasch-, unweighted mean, and weighted mean scores, and this was the case for all three groups under consideration. Additionally, we performed a factor analysis on all three scoring methods. For each of the three groups a single factor emerged explaining variance in excess of 96%, on which all three scoring methods loaded very highly (table 2.11) As a result of these investigations, we have established that

[5] Values in excess of 0.7 are normally considered to signal very strong reliability.

all three aggregation methods yielded comparable results and could be used interchangeably for analysis purposes.

Table 2.11 Factor loadings on satisfaction items

	Variance accounted for	Loading on Factor 1	Loading on Factor 2	Loading on Factor 3
Teaching	0.964	0.973	0.993	0.979
Assessment	0.978	0.983	0.994	0.989
General	0.989	0.989	0.997	0.997

In this section we condensed the satisfaction questions in the Student Outcome Survey into three coherent components. Results of the principal component analysis showed this partition was statistically valid. We also created summary measures that encapsulated the three main themes of student satisfaction to aid future research and reporting. To achieve this, three different quantitative methods were devised, evaluated and compared. While the three methods each employed a distinct scoring technique, as far as the measurement of the core outcome for each category was concerned, the statistical outcomes differed very little.

So which method was the most appropriate one to use in subsequent analyses? Given that all three methods yielded very similar results it seemed reasonable to rely on the measure that was most easily created. In our case, this would have been to calculate unweighted mean scores for the three components that we identified earlier. However, as we were planning to subsequently utilise the resulting summary scores in our analysis of performance indicators it appeared more prudent to take advantage of the composite scores created by Rasch analysis. These scores exhibited a greater variance and also better fitted the requirement of a continuous, true interval variable for subsequent regression analyses.

2.1.11 Selection bias in the Student Outcomes Survey

The overall response rates for the SOS in 2011 were 42.2% for graduates and 32% for module completers. Response rates for the government providers

under consideration in this study were slightly higher with 45.6% for graduates and 32.2% for module completers. There was considerable variation in response rates across the states and territories, with response rates for module completers exhibiting the largest variation, ranging from 22.9% in the Northern Territory to 38.4% in the Australian Capital Territory (table 2.12; NCVER 2011e).

Table 2.12 Response rates for graduates and module completers in the 2011 SOS

State/territory	Graduates	Potential module completers
	%	%
New South Wales	44.1	32.8
Victoria	42.4	34.3
Queensland	40.4	29.0
South Australia	42.8	37.7
Western Australia	41.9	30.0
Tasmania	46.5	33.9
Northern Territory	37.7	22.9
Australian Capital Territory	39.1	38.4
Total	**42.2**	**32.0**

While randomly distributed non-response would be only having an impact on statistical power and confidence intervals, non-random response patterns may lead to non-response bias and thus contribute significantly to systematic error in the survey. The Student Outcome Survey addressed this issue somewhat by applying a post-stratification scheme which weighed responses to population benchmarks of students who completed their training in 2011, the target population of the survey. However, those benchmarks, namely sex, age, study location, field of education and group (graduate and module completer) covered only a small spectrum of possible sources of response bias. Moreover, the application of this technique assumed that there were coherent response patterns that, for instance, separated responding males from non-responding males, which was not necessarily the case. It could thus be speculated that selection bias still existed even after post-stratification.

While there are some well-known approaches in survey statistics that deal with this problem, such as raking (see Deville et al., 1993) or propensity scoring (see Leow et al. 2004), in our analysis we employed the two-step consistent estimator method developed by Heckman (1979).

2.1.12 The two-step Heckman method in models developed in this study

Heckman developed his method by postulating that sample selection bias may be seen as specification error amounting to 'omitted variable' bias. If the values for this omitted variable can be estimated they could then be used as a predictor in the equation that is used to estimate the behavioural function of interest (Heckman 1979). Conceptually, the Heckman methodology can be seen as a model of two separate equations. In the context of this study, there would be the equation of actual interest (for instance, student satisfaction) and the selection equation, which ascertains whether there was a response for the satisfaction items. The corresponding models in our case are the equation of our substantive interest, given as:

$$satisfaction = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + u_1 \qquad (3)$$

and the selection equation:

$$selection = \gamma_0 + \sum_{i=1}^{n} \gamma_i X_i + u_2 \qquad (4)$$

where u_1 and u_2 have the correlation φ. Xi is a vector of explanatory variables including SES, age, hours enrolled, field of education, sex, funding source, remoteness region, student type (module completer/graduate), qualification level, Indigenous indicator and institute.

There are two variants of Heckman's methodology which yield almost identical results. The first method brings together the two components of the Heckman procedure and estimates the results employing maximum likelihood estimation. This method is computationally intensive and does often end in convergence problems for large data sets. The second method is the two-step (also called Heckit-) method. In this method the estimation of selection bias is

separated into two discrete steps. The first is the estimation of the values of the 'omitted variable'. This involves the calculation of the inverse Mills ratio (IMR). The IMR is derived via a probit model estimating the probability of students responding to the survey and represents essentially the ratio of the probability density function over the cumulative distribution function of the linear predictor for response. The formula for the IMR is thus:

$$IMR = \frac{\Phi_{prob}(Z)}{\Phi_{dist}(Z)} \quad (5)$$

where Z is the linear predictor for the probability of receiving a response. When the IMR is entered as a fixed effect into a regression model, its significance indicates existing selection bias with respect to the outcome variable, while non-significance indicates the absence of response bias and the IMR variable can then be removed from the model.

There are several reasons for using the Heckman technique in this research. The first is that we have access to significant demographic and educational background data in responding and non-respondent records, which allows us to estimate the probability of receiving a response from every sampled student. The second reason is that this method enabled us not only to determine whether or not selection bias is present in our outcome variables, but also to estimate the extent of that potential bias. We will be using the two-step method of the Heckman technique, as the one-step method leads to convergence problems due to the large size (ca. 356,000 observations) of the SOS data set that includes non-respondents. Puhani (2002) reported that the two-step estimator was also preferable in terms of dealing with multicollinearity, while producing reasonable results.

In this research we estimated two separate IMRs. The reason for this was that we have two different item response patterns in the SOS. While those who respond to the survey usually answered every item in the questionnaire for

which they were eligible, this was not the case for the item of post-training salary. The salary question received only about a 30% response from those who returned the questionnaire. We therefore concluded that there was a separate response profile for the salary item and estimated a distinct response probability that was then used solely in the salary model later in the analysis.

The independent variables used to estimate individual response probabilities were a socioeconomic background proxy (for example, SEIFA percentile[6], numerical), age (numerical), number of hours enrolled (four categories, <50, 50—199, 200—540, >540), field of education (69 categories, as per Australian Standard Classification of Education (ASCED), ABS (2013a), sex (two categories, male and female), funding source (five categories; Commonwealth, Commonwealth specific purpose, state, fee paying, and revenue from other provider), remoteness (six categories as per ABS ARIA definition [ABS, 2013b]), student type (two categories; graduate or module completer), qualification level (15 categories; as per Australian Qualification Framework [AQF ,2013]), Indigenous status (three categories; Indigenous, non-Indigenous, unknown) and institution attended (categorical; 96 institutions covered in the SOS). All categorical variables were dummy coded. Model fit and analysis of effects (estimated using Type 3 sum-of-squares) are displayed in table 2.13. As can be seen, all independent variables are significant in predicting the outcome of 'response'. Very high Chi-squared values for age, remoteness and institute indicate that these variables were particularly useful.

The results from the analysis of type 3 effects for the separate 'salary response' model looks very similar, with the exception of the non-significance of 'enrolment hours' variable (see table 2.13).

[6] Socio Economic Indexes For Areas (SEIFA) is a product developed by the Australian Bureau of Statistics that ranks areas in Australia according to relative socioeconomic advantage and disadvantage. The indexes are based on information from the five-yearly Census. For more information see ABS (2013).

Table 2.13 Type 3 effects probit modelling response likelihood

Effect		Overall		Salary only	
	DF	Chi-Sq	P>ChiSq	Chi-Sq	P>ChiSq
Percentile	1	39.3	<.0001	65.6	<.0001
Age	1	5418.0	<.0001	2494.2	<.0001
Hours enrolled	3	402.3	<.0001	1.6	0.6584
Field of education (4 digit)	68	1415.9	<.0001	2714.2	<.0001
Sex	1	1234.1	<.0001	144.7	<.0001
Funding source	4	813.8	<.0001	833.6	<.0001
Remoteness region	5	3012.9	<.0001	654.2	<.0001
Student type	1	66.9	<.0001	1253.1	<.0001
Qualification level	14	1643.5	<.0001	1973.5	<.0001
Indigenous	2	172.6	<.0001	100.2	<.0001
Institute	95	3313.4	<.0001	1533.3	<.0001
Wald ChiSq			<.0001		<0.001
Pseudo RSq			0.12		0.13

In the case of these two ‘response models’ we were of course only moderately interested in coefficients and model statistics themselves, but much more in the use of these to estimate individual linear response probabilities. These linear response probabilities were then used to calculate the inverse Mills ratios for every sampled student. In addition to the three Rasch component variables from the earlier section, we added two more new variables containing IMR and IMR_{Salary} to our data set for subsequent use in our models.

2.1.13 Statistical models for institutional performance assessment

The data setup for this analysis had a nested structure. For example, the students were nested within each school, the unemployment rate nested within a given labour force region, or the income nested within the prevailing income conditions within a postcode area. In such a setting there is variability between individual institutes, as well as variability between students nested within these institutes. To deal with this nested, multi-level structure of the data, mixed models (also known as hierarchical or multi-level models) are employed (Dai & Rocke, 2006). Mixed models are models that deal with data which vary at more than one level and contain both fixed and random effects (Raudenbush & Bryk, 2002). The main method used for the comparison of institutional performance indicators was thus multi-level regression modelling.

In this analysis of institutional performance we considered two types of outcome variables: employment status, perception of achievement, and willingness to recommend the institution of training were all dichotomous outcomes, while the various types of student satisfaction and salary after training were continuous outcomes.

Using employment after training as an example, a conventional logistic regression model (see Hosmer & Lemeshow, 2000) would take the general form:

$$logit(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha + \beta x_{ij} \qquad (6)$$

with π being the probability of being employed and subscripts *i* and *j* indicating the two levels of students and their training institutions. The probability function for the logit model in (4) is:

$$\pi_{ij} = \frac{exp(\alpha+\beta x_{ij})}{1+exp(a+\beta x_{ij})} \qquad (7)$$

To extend the logit model (4) concept to account for the random effect an additional element needs to be added into the equation:

$$logit(x_{ij}) = \alpha + u_j + \beta x_{ij} \qquad (8)$$

and

$$u_j = a_j - a \qquad (8a)$$

where a_j is a random intercept as a linear combination of *a* (the grand mean) and u_j (the deviation from that grand mean).

The logit model in (6) is a mixed model as it accommodates the fixed *(α, β)* and random effects *(u_j).* It is also a logistic model, as the link function is logit and thus a subclass of a generalised linear mixed model. We used this model to estimate institutional effects where our dependent variable was dichotomous, for example, employment, willingness to recommend their institution of

training and sense of achievement of goals. Models estimating the various types of student satisfaction had a semi-continuous dependent variable and therefore employed ordinary mixed-effect models. These were methodologically similar to the logistic models described above but did not use a logit link function.

In all the performance indicator models that have been developed throughout this paper we performed the analysis without the use of survey weights. There were two major reasons for this: The first reason followed from the rationale advanced by Winship and Radbill (1994) who argued that in cases where survey weights were solely a function of the independent variables in a model, the unweighted regression estimates were preferable as they were unbiased, consistent and had smaller standard errors than weighted regression estimates. The second reason related to the actual design of the SOS, and to our intention to address the suspected selection bias in the survey. SOS selection weights from the outset did not exist to adjust for the different probabilities of selection. The survey employed a systematic sampling approach across the entire population in which every potential respondent had the same probability of being selected. The survey weights were created by post-stratification weighting in an effort to address non-response (NCVER, 2011e). Since we did endeavour to use Heckman selection models as a means to account for possible respondent bias we omitted the utilization of post stratification weights.

2.1.14 Individual performance indicators

In this study we aimed to assess the institutional effect on eight different outcomes, namely student satisfaction with teaching, assessment, general learning, overall satisfaction, employment outcomes, salary outcomes, perception of achievement and willingness to recommend the institution of training. In order to do this we used two closely-related methodologies, mixed

effect modelling for continuous outcomes and mixed effect modelling for dichotomous outcomes. While we developed eight different models for eight different outcomes, all models shared some covariates in common. These covariates were the ones that were meant to enable us to evaluate TAFE performance on an 'even footing'. For example, these covariates were used to adjust our models for the specific demographic and educational characteristics of individual institutions. To illustrate, consider an institute that has a very high proportion of students who, due to their chosen course, will end up in the mining industry. If employment opportunities in the mining industry are particularly favourable this would mean that the institute would have a high employment performance score. This, however, may be mostly due to the courses its students pick and not due to institutional characteristics itself. We therefore added the dummy-coded 'field of education' into our model(s), in an effort to adjust for the differences between institutes in terms of the field of study chosen.

The variables common to all our models were

- Age
- Sex
- Qualification level
- Field of education
- Group status
- Inverse Mills ratio

Age was entered into models as a continuous variable as calculated from the birth date in the sampling records. Sex was coded as a dichotomous variable, where 1 indicated male and 0 indicated female. The qualification level variable was a categorical variable containing eight levels of TAFE qualifications according to the AVETMISS standard (AVETMISS, 2011). The qualification levels were 'Subject only enrolment', 'Statement of attainment',

'Other', 'Certificate I', 'Certificate II', 'Certificate III', 'Certificate IV', and 'Diplomas and above'. The 'field of education' variable was a categorical variable indicating 13 areas of study. These were 'Natural and physical sciences', 'Information technology', 'Engineering and related technologies', 'Architecture and building', 'Agriculture, environmental and related studies', 'Health', 'Education', 'Management and commerce', 'Society and culture', 'Creative arts', 'Food, hospitality and personal services', 'Mixed field programs', and 'Subject only enrolment'. Group status was a categorical variable designating whether the student was a module completer or a graduate. We shall refer to the covariates described above as our *standard covariates*. In addition, the inverse Mills ratio was a continuous variable relating to the statistical probability of the individual student responding to the survey. The model of every performance indicator was first run with the IMR variable entered as a fixed effect. Where the IMR contribution to the model was significant it was retained in the model, otherwise the IMR was discarded and the model re-estimated without the IMR.

2.1.15 Findings: employment outcomes

The first performance indicator that we developed in this study dealt with whether there is an institutional effect on employment outcomes. We created an employment outcome indicator from the SOS 'Labour force status after training' variable. This was a dichotomous variable denoting the categories 'employed' or 'not employed' after training. We also added a number of covariates into our employment outcome model. These covariates were divided into fixed and random effects.

Fixed effect covariates in this model were our standard covariates, the unemployment rate, and the labour force status before training. It had been shown that the most important predictor of post-study employment outcomes was employment status prior to the training (Karmel & Fieger, 2012). The

inclusion of the labour force status prior to training thus provided an additional means to adjust for varying characteristics between institutions. In this model of institutional effects on employment outcomes, we added the institution of the respondent and unemployment rate in the local area of individual students as a random effect. It seemed reasonable to assume that it is harder to obtain employment in areas with a high unemployment rate than it is if the student lived in a region of virtually full employment. The rationale for incorporating this random effect into our model was the same as outlined before: to adjust for local variations and to more clearly delineate the effect the training institutions had on employment outcomes. In this and all following mixed models, categorical fixed effect variables were dummy coded (and their resulting estimates thus representing contrasts with respect to the reference category), whereas the main random effect variable (for example, the institution) was effect coded in order to enable us to estimate contrasts with respect to the overall mean.

Table 2.14 displays all type 3 fixed effects in the employment model. Our first consideration was whether the IMR had a significant effect. The 0.25 Pr>F value indicates that the IMR did not significantly differ from zero and that therefore selection bias was not an issue for the employment outcome. We could therefore base our analysis on the model which did not include the IMR (and all following tables stem from the model omitting the IMR).

Table 2.14 Type 3 effects on employment outcome

Variable		Basic		Basic + Mills	
	DF	F	Pr > F	F	Pr > F
Field of education (2-digit)	11	82.9	<.001	82.6	<.001
Employed before	3	3463.4	<.001	3394.4	<.001
Age	1	4.0	0.045	22.0	<.001
Sex	1	9.0	0.003	10.4	0.001
Qualification level	6	99.4	<.001	92.8	<.001
Unemployment	1	2.1	0.150	0.0	0.977
Graduate/module completer	1	101.7	<.001	92.9	<.001
Inverse Mills ratio	1	-	-	1.3	0.250

All fixed effects except for age were significant and had thus predictive value for the employment outcome. The covariance parameter for the random effect of the prevailing unemployment rate in the local area of the respondent (table 2.15) was significantly different from zero. This indicated that the local labour force situation played a role in labour market outcomes for graduates and module completers. It also adjusted the mixed effect model for these parameters. However, the non-significance of the fixed effect of the unemployment rate (table 2.15) pointed to a limited overall impact of the unemployment rate on employment outcomes.

Table 2.15 Covariance parameter estimates – employment model

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	0.033	0.009	3.730	<.001
Unemployment rate		0.001	<.001	2.943	0.003

The covariance parameter estimate for the intercept of the teaching institution was significant. This means that, after adjusting for the aforementioned covariates, there were institutional differences in employment outcomes, and that these institutional effects were measurable.

In order to estimate the overall effect of TAFE institutes on employment outcomes it was necessary to calculate the intraclass correlation (ICC). The ICC is a measure of the proportion of the variance in post-training employment explained by the grouping of students within the various TAFE institutes. In other words, it can be seen as the proportion of the total variance in post-training employment which is explained by the difference between institutes. In a mixed effects linear model such as the one above, the ICC shows the total variance that is accounted for by the clustering. It can also be conceptualised as the correlation of observations within the same cluster. In practical terms, the ICC is a measure of how much of the variation in employment outcomes is due to the sample being clustered into different institutes. To facilitate the calculation of the ICC we needed to separately estimate an intercept-only

model with all predictor variables removed except for the random effect of institutions (Baecke & Van den Poel, 2012). We could then derive the ICC by dividing the estimate of the covariance between institutions by the covariance between institutions, and residuals:

$$ICC = \frac{\sigma^2{}_{Institutes}}{(\sigma^2{}_{Institutes} + \sigma^2{}_{Residual})} \quad (9)$$

The covariance term of the residual in a model with a logistic distribution was approximated (as proposed by Hedecker, 2003) by:

$$\sigma^2{}_{residual} = \frac{\pi^2}{3} \quad (10)$$

Table 2.16 displays the estimate of the variance of the intercept-only employment model. The ICC for our post-training employment model is 0.0365. This indicates that 3.65% of the total variation in employment can be attributed to differences between TAFE institutes. It should thus be kept in mind that the majority of the variance in employment arises from the variation between individual students rather than between institutes.

Table 2.16 Covariance parameter estimates intercept-only employment model

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	0.125	0.029	4.250	<.001
Intraclass correlation		0.036			

Results for individual institutions can be found in table 2.17. Green shading signifies that the institutional performance with respect to post-training employment featured significantly above the mean for all providers. Red shading indicates below-average performance. In the table, we can see that a number of institutes (4, 8, 10, 29, 37, 38, 45, 50, and 71) performed significantly below the average of TAFE institutes and institutes 19, 30, and 70 performed significantly above the average TAFE institute with respect to employment outcomes. This was true after institutional demographics, course

profiles, and predominant employment conditions in the residential area of students had been taken into account.

Table 2.17 Intercept estimates for institutional employment effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	-0.135	0.091	-1.50	0.135	36	0.025	0.104	0.24	0.808
2	-0.059	0.105	-0.56	0.573	37	-0.244	0.116	-2.11	0.035
3	-0.010	0.129	-0.08	0.936	38	-0.254	0.089	-2.87	0.004
4	-0.275	0.103	-2.66	0.008	40	-0.057	0.144	-0.40	0.691
5	-0.114	0.085	-1.33	0.183	43	-0.032	0.090	-0.35	0.724
6	-0.119	0.100	-1.19	0.233	44	0.144	0.092	1.58	0.115
7	0.032	0.117	0.27	0.784	45	-0.306	0.118	-2.59	0.010
8	-0.404	0.085	-4.75	<.001	46	0.164	0.119	1.38	0.169
10	-0.182	0.074	-2.45	0.014	47	0.000	0.101	0	0.999
11	-0.129	0.112	-1.16	0.247	48	-0.011	0.079	-0.14	0.891
12	-0.096	0.089	-1.08	0.280	49	0.074	0.109	0.67	0.500
13	0.144	0.082	1.75	0.081	50	-0.226	0.088	-2.58	0.010
14	0.002	0.112	0.02	0.986	51	-0.056	0.072	-0.78	0.435
15	0.165	0.092	1.79	0.073	52	-0.044	0.120	-0.36	0.716
16	0.086	0.110	0.79	0.432	53	0.005	0.121	0.04	0.966
17	-0.002	0.117	-0.02	0.985	55	0.130	0.125	1.04	0.300
18	-0.010	0.101	-0.10	0.921	56	-0.012	0.133	-0.09	0.928
19	0.242	0.113	2.15	0.032	57	0.070	0.130	0.54	0.591
20	-0.082	0.079	-1.04	0.298	58	0.204	0.129	1.58	0.115
22	0.014	0.090	0.16	0.873	60	0.088	0.137	0.64	0.521
23	-0.104	0.090	-1.16	0.245	61	0.180	0.158	1.14	0.254
24	-0.166	0.093	-1.78	0.074	64	0.100	0.091	1.10	0.273
25	-0.006	0.118	-0.05	0.961	65	-0.010	0.088	-0.11	0.909
26	-0.096	0.127	-0.75	0.451	66	-0.037	0.084	-0.44	0.662
27	0.045	0.077	0.58	0.563	70	0.329	0.101	3.28	0.001
28	0.141	0.119	1.19	0.234	71	-0.230	0.093	-2.48	0.013
29	-0.195	0.088	-2.22	0.027	74	-0.040	0.162	-0.25	0.805
30	0.267	0.107	2.51	0.012	75	0.117	0.095	1.22	0.221
31	0.154	0.110	1.40	0.162	77	0.189	0.101	1.88	0.060
32	0.096	0.127	0.75	0.451	78	-0.077	0.175	-0.44	0.661
33	0.031	0.108	0.29	0.772	109	-0.012	0.168	-0.07	0.943
34	0.019	0.076	0.26	0.797	110	0.139	0.128	1.08	0.278
35	0.134	0.109	1.23	0.218					

Notes: = Institutions performing significantly above the mean.
= Institutions performing significantly below the mean.

A more visual presentation of the above table that is easily interpretable can be found in the caterpillar plot in figure 2.3. Here we have sorted all institutes by estimate from smallest to largest and added error bars (95% confidence interval) to account for institutional sample size and variability.

Figure 2.3 Intercept estimates for institutional employment effects

It can clearly be seen that there are a number of institutes which performed significantly below average (i.e. error bars do not cross the 0 line), and a lower number of institutes performed above average.

2.1.16 Wage outcomes

The second performance indicator we developed and analysed was the annual salary/wages of students after their training and whether there was an institutional effect on these salaries. Salary data were available only for a subset of SOS students, as many respondents chose not to reveal their income details. Despite this, there was still a significant pool of around 25,000 respondents so that statistically valid inferences could be drawn from the data.

Income data for individual students was derived by using the midpoint of a number of possible annual income ranges that are offered in the SOS. While this was not an ideal method (for instance, it somewhat eliminated exceptionally large incomes) it still provided an acceptable proxy for individual incomes, especially in a comparative sense, as the same method of derivation was applied to every survey respondent. Fixed effect covariates entered into this model included the standard covariates, students' labour force status before their training, occupation after training and the industry after training. We

hypothesised that those who had been employed prior to their training were able to command higher salaries and that this justified including 'labour force status before training' as an independent variable in the model. Similarly, it was assumed that occupation and industry of the respondent impinged on income levels and it was therefore necessary to adjust for them in order to account for differing institutional profiles with respect to these covariates.

In this model of institutional effects on post-training incomes, we added the institution and the deviation of the local area of the respondent from the average national annual income as a random effect. We hypothesized that the post-training incomes that can be achieved were influenced by prevailing average incomes in the local area of the respondents and it was likely to be worthwhile to adjust for this possibility.

The significance parameters for the type 3 fixed effects can be found in table 2.18. The effect of the IMR_{Salary} was highly significant, indicating that there was significant selection bias in our income model. It was therefore prudent to retain the IMR_{Salary}, and all further analyses in this section were performed using the model which included the IMR_{Salary}. All other fixed effects were significant.

Table 2.18 Type 3 effects on salary outcome

Variable		Basic		Basic + Mills	
	DF	F	Pr > F	F	Pr > F
Field of education (2-digit)	11	9.3	<.001	4.6	<.001
Employed before	4	129.9	<.001	81.4	<.001
Age	1	161.3	<.001	894.0	<.001
Sex	1	470.7	<.001	300.0	<.001
Qualification level	6	77.1	<.001	27.8	<.001
Occupation after training	51	62.1	<.001	54.4	<.001
Industry after training	86	32.5	<.001	30.7	<.001
Annual income deviation	1	141.8	<.001	144.8	<.001
Graduate/module completer	1	2.6	0.1	8.6	0.003
Inverse Mills ratio	1	-	-	72.8	<.001

The covariance parameter for the random effect of the local income variable was significantly different from zero (table 2.19). This indicated that the local prevailing income situation played a role in achieved individual post-training incomes. The covariance parameter estimate for the intercept of the teaching institution was significant. Therefore, after adjusting for the aforementioned covariates, there was a measurable institutional effect on post-training incomes.

Table 2.19 Covariance parameter estimates – salary model

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	5346164	854740	6.25	<.001
Annual income	Postcode	0.095	0.025	3.74	<.001
Residual		347430000	3039702	114.30	<.001

Table 2.20 shows the covariance estimates for the intercept-only salary model which were then used to derive the ICC. The ICC in our salary model was 0.0422. This indicated that 4.2% of the total variation in salaries could be attributed to differences between TAFE institutes.

Table 2.20 Covariance parameter estimates – intercept-only salary model

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	25553033	3545817	7.21	<.001
Residual		579570000	4928112	117.61	<.001
Intraclass correlation			0.042		

Individual institutions' results are shown in the following table 2.21. The institutional estimates can be interpreted as the institutional difference from the overall income mean, after all the covariates have been taken into account.

Table 2.21 Intercept estimates for institutional salary effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	211	783.8	0.27	0.788	36	-665	886.5	-0.75	0.453
2	215	1045.3	0.21	0.837	37	-1711	1090.0	-1.57	0.116
3	-1476	1210.3	-1.22	0.223	38	-1834	968.6	-1.89	0.058
4	-1965	1022.3	-1.92	0.055	40	31	1827.8	0.02	0.987
5	-846	917.1	-0.92	0.356	43	-19	991.5	-0.02	0.985
6	-1555	1041.0	-1.49	0.135	44	-758	832.1	-0.91	0.362
7	-2188	1060.6	-2.06	0.039	45	515	1098.7	0.47	0.639
8	45	798.3	0.06	0.955	46	200	1151.9	0.17	0.862
10	-229	747.3	-0.31	0.759	47	-448	800.5	-0.56	0.576
11	1040	978.3	1.06	0.288	48	2514	685.4	3.67	<.001
12	796	831.5	0.96	0.339	49	-490	1001.7	-0.49	0.625
13	-1833	708.6	-2.59	0.010	50	4934	759.4	6.5	<.001
14	-2966	1133.9	-2.62	0.009	51	1429	699.7	2.04	0.041
15	-2200	853.1	-2.58	0.010	52	4117	1082.6	3.8	<.001
16	-1571	932.9	-1.68	0.092	53	1006	1142.9	0.88	0.379
17	-3035	1109.2	-2.74	0.006	55	1860	1045.1	1.78	0.075
18	-2049	832.6	-2.46	0.014	56	7302	1296.0	5.63	<.001
19	-1277	1014.3	-1.26	0.208	57	2194	1211.9	1.81	0.070
20	-2713	750.5	-3.62	<.001	58	9770	1211.3	8.07	<.001
22	-4489	837.6	-5.36	<.001	60	2979	1375.9	2.16	0.030
23	-1953	857.4	-2.28	0.023	61	608	2149.5	0.28	0.777
24	-840	992.8	-0.85	0.398	64	-1494	738.1	-2.02	0.043
25	-2210	1061.9	-2.08	0.037	65	-849	818.3	-1.04	0.300
26	-2999	1318.0	-2.28	0.023	66	-3362	788.9	-4.26	<.001
27	3531	775.7	4.55	<.001	70	1379	698.1	1.98	0.048
28	1177	1047.1	1.12	0.261	71	-2168	1121.8	-1.93	0.053
29	275	893.6	0.31	0.759	74	831	1927.1	0.43	0.666
30	-1940	1084.5	-1.79	0.074	75	2309	882.3	2.62	0.009
31	-1958	930.6	-2.1	0.035	77	1411	914.9	1.54	0.123
32	-1024	1118.3	-0.92	0.360	78	-3339	2156.0	-1.55	0.122
33	-1133	1141.7	-0.99	0.321	109	258	2216.1	0.12	0.907
34	2245	782.1	2.87	0.004	110	806	1295.2	0.62	0.534
35	5599	895.2	6.25	<.001					

Notes: = Institutions performing significantly above the mean.
= Institutions performing significantly below the mean.

Green shading signifies that the institutional performance in respect to post-training incomes features significantly above the mean of all providers. Red shading indicates below-average performance. A graphical presentation of the sorted performance data is provided in figure 2.4. It can be seen that there was a steep rise at the top end of TAFE institutes, with four institutes not only performing well above the mean, but also significantly better than most other institutes.

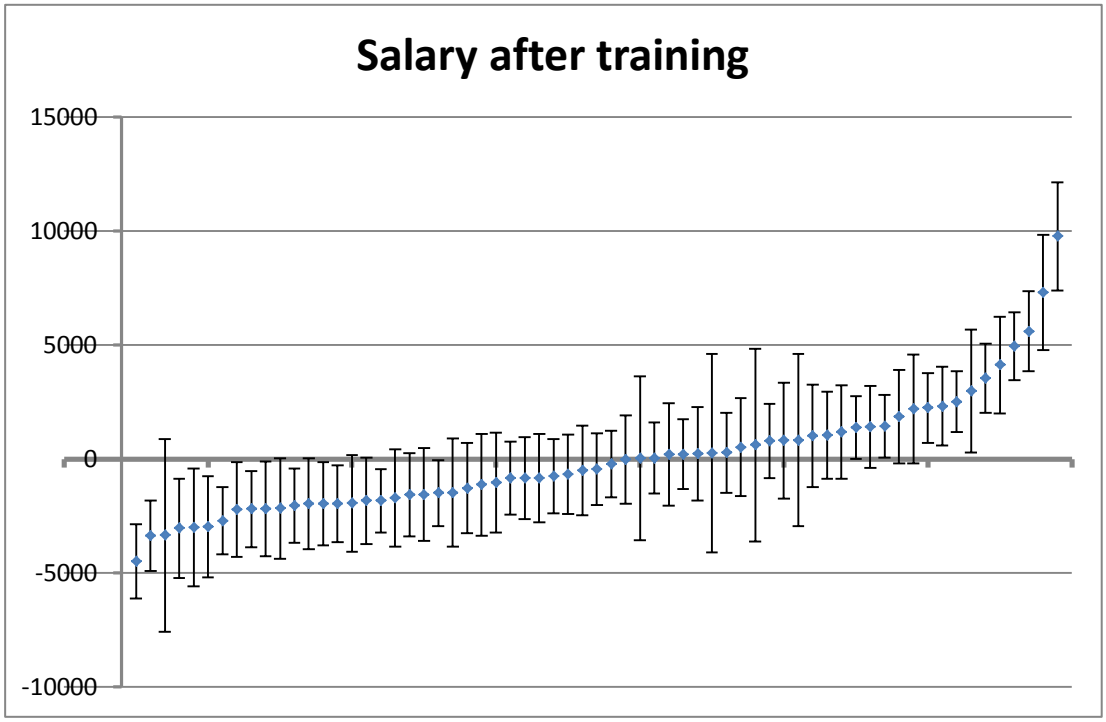

Figure 2.4 Intercept estimates for institutional salary effects

The extent of the selection bias is also of interest in the salary model since the inverse Mills ratio was significant. Table 2.22 contains the complete solution for all fixed effects under the simple salary model and the salary model adjusted for selection effects. We can see that the estimate for the IMR is -5,177 (also known as lambda). Heckman (1979) postulated that the relationship

$$\rho = \frac{\lambda}{\sigma} \qquad (11)$$

holds, where σ is the standard error of the residuals and ρ is the correlation coefficient between the unobservable influences that determine response to the salary question in the SOS and the unobservable influences that determine the salary. As lambda in our salary model was negatively signed it can be concluded that the correlation between the selection equation and salary equation was negative, which in turn suggested that response to the salary item in the survey was associated with lower salaries.

Table 2.22 Difference in salary estimates: adjusted v non-adjusted for selection bias

Effect	Heckman		No Heckman		Difference	t	Pr > t
	Estimate	Std err	Estimate	Std err	Estimate		
Intercept	48361	3642	50163	3462	1802	0.36	0.719
IMRsalary	-5177	599	.	.	.	.	.
Age	345	12	42	3	-303	-25.22	<0.001
Income deviation	0	0	0	0	0	0.07	0.944
Sex							
Female	-5480	316	-6933	320	-1453	-3.23	0.001
Male	0	.	0	.	0	.	.
Group							
Graduate	-1326	451	-714	440	612	0.97	0.332
Module completer	0	.	0	.	0	.	.
Employed before training							
Employed	2822	2778	2364	2849	-458	-0.12	0.904
Unemployed	-5498	2825	-7803	2897	-2305	-0.57	0.569
Not in labour force	-3777	2836	-6911	2908	-3134	-0.77	0.441
Not employed	-4012	3938	-5149	4039	-1137	-0.2	0.841
Not stated	0	.	0	.	0	.	.
Field of education							
Natural & Phys. Sciences	-4135	2011	-2170	2060	1964	0.68	0.497
Information Technology	68	936	2111	938	2042	1.54	0.124
Engineering & Related	892	665	2895	648	2003	2.16	0.031
Architecture & Building	-1253	770	318	768	1571	1.44	0.15
Agriculture & Related	-167	812	2430	806	2597	2.27	0.023
Health	-937	734	1307	727	2244	2.17	0.03
Education	463	745	2884	738	2421	2.31	0.021
Management & Commerce	-1164	675	818	658	1982	2.1	0.036
Society & Culture	-1912	724	-623	734	1288	1.25	0.211
Creative Arts	-2547	1076	-3935	1100	-1389	-0.9	0.368
Food, Hosp. & Pers. Serv.	485	781	954	795	469	0.42	0.674
Mixed Field Programs	0	.	0	.	0	.	.
Qualification level							
Diploma & above	255	752	-463	770	-718	-0.67	0.503
Certificate IV	-362	725	-644	743	-282	-0.27	0.787
Certificate III	-2624	716	-5255	724	-2631	-2.58	0.01
Certificate II	-5767	769	-8784	761	-3017	-2.79	0.005
Certificate I	-1802	1115	-5311	1088	-3509	-2.25	0.024
Other	966	654	1322	670	356	0.38	0.704
Statement of attainment	0	.	0	.	0	.	.
Occupation							
51 categories			no significant differences				
Industry							
87 categories			no significant differences				

Note: Shaded predictors are significant.

Another item of interest was to ascertain which specific variables contributed to biased estimates in the model without adjustment for selection bias. Here it

was necessary to compare estimates for the same variables with and without the inclusion of the IMR. The data are presented in table 2.22. Sales et al. (2004) suggested looking for a change of >10% between adjusted and non-adjusted estimates; however, they conceded that this criterion was arbitrary. In this study we applied a two-sample t-test, with assumed similar variances between the two estimates. We calculated the t statistic as follows:

$$t = \frac{Estimate_{IMR} - Estimate}{\sqrt{StdErr_{IMR}{}^2 + StdErr^2}} \qquad (12)$$

Table 2.22 thus contains the t-statistic as well the probability parameters. Only two independent variables were considered to be different: The age parameter differed from $345 in the adjusted model to $42 in the unadjusted model (p <.001). This suggested that the selection pattern in the unadjusted model was biasing salaries downward. The parameter for female sex was significantly different between the adjusted model (-$5,480) and unadjusted model (-$6,933), implying that the model unadjusted for selection bias was yielding unrealistically low estimates. There were also several sub-categories of the 'qualification level' variables, such as certificate 1, 2, and 3, and 'field of education' variables, such as engineering, agriculture, health, education, and management, that generated significantly different results in unadjusted and adjusted models.

2.1.17 Student satisfaction with teaching, assessment and learning

Student satisfaction was an obvious performance indicator, as it directly reflected a student's subjective appreciation for the training they had received. To recap, the Student Outcomes Survey captured student satisfaction by asking a battery of 19 questions evaluating students' satisfaction with 19 different aspects of their training. These 19 questions were grouped under three main themes: teaching, assessment, and general skills and learning experiences and each contained responses on a 5-step Likert-style scale, with '1' indicating strong dissatisfaction, and '5' indicating

strong satisfaction. We previously validated the internal consistency of the three main themes and created a summary measure for each theme as described in section 1.5. For the analysis of institutional effects on satisfaction we used the three Rasch composite satisfaction scores created and described in that section. Due to the identical modelling approach in all three sub-types of student satisfaction we will deal with them collectively in this section.

In addition to the 19 themed satisfaction questions (and thus the three themes of teaching, assessment and learning), the Student Outcomes Survey contained a summary question asking for students' overall satisfaction with their training. As this variable had not been created via a Rasch model and therefore had different properties relative to the three component variables, we dealt with it separately in the following section. There were thus three models constructed to examine the effect of TAFE institutes on student satisfaction, dealing with teaching, assessment and learning. Independent variables in all three models were our standard covariates.

Table 2.23 Type 3 effects on satisfaction outcomes

Variable		Teaching				Assessment				Learning			
		Basic		Basic + Mills		Basic		Basic + Mills		Basic		Basic + Mills	
	DF	*F*	*P > F*	*F*	*P > F*	*F*	*P > F*	*F*	*P > F*	*F*	*Pr > F*	*F*	*P > F*
Field of educ.	11	36.8	<.001	32.8	<.001	25.2	<.001	23.5	<.001	35.9	<.001	36.8	<.001
Age	1	14.6	<.001	121.2	<.001	0.0	0.847	11.2	<.001	70.4	<.001	163.9	<.001
Sex	1	0.3	0.563	2.3	0.139	5.7	0.020	6.4	0.014	50.1	<.001	35.8	<.001
Qual level	6	76.4	<.001	55.3	<.001	21.4	<.001	20.5	<.001	19.4	<.001	19.3	<.001
Graduate/MC	1	448.2	<.001	481.0	<.001	514.0	<.001	503.2	<.001	999.2	<.001	921.0	<.001
IMR	1	-	-	19.7	<.001	-	-	0.6	0.433	-	-	0.8	0.360

In table 2.23 we can see that only the satisfaction with teaching model component exhibited selection bias, as indicated by the significant IMR. In the case of this component of satisfaction we therefore proceeded with the model that included the IMR. All other standard covariates were significant, except for sex in the satisfaction with teaching model. In table 2.24 we report the covariance parameter estimates for all three models.

Table 2.24 Covariance parameter estimates – satisfaction models

		Teaching		Assessment		Learning	
Parameter	Subject	Estimate	Pr > t	Estimate	Pr > t	Estimate	Pr > t
Intercept	Institute	0.060	<.001	0.045	<.001	0.043	<.001
Residual		7.587	<.001	7.782	<.001	7.997	<.001

We found that there were significant institutional effects on all three types of satisfaction. Intercept-only satisfaction models along with calculated ICCs are shown in table 2.25.

Table 2.25 Covariance parameter estimates – intercept only

		Teaching		Assessment		Learning	
Parameter	Subject	Estimate	Pr > t	Estimate	Pr > t	Estimate	Pr > t
Intercept	Institute	0.085	<.001	0.045	<.001	0.039	<.001
Residual		7.744	<.001	7.933	<.001	8.349	<.001
Intra class correlation		0.011		0.006		0.005	

It is clear that the variation (from 0.5 to 1.1%) in the three sub-types of satisfaction explaining the difference between TAFE institutes is much lower than in the employment and salary models.

Institutional intercept estimates for satisfaction with teaching, assessment, learning and overall satisfaction are displayed below in tables 2.26 (a—c). Institutes that performed significantly above the mean of institutes have been shaded in light green and those that performed significantly below the mean were coloured in light red.

Table 2.26a Intercept estimates for institutional teaching satisfaction effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	0.146	0.076	1.91	0.056	36	0.089	0.087	1.02	0.309
2	0.260	0.093	2.78	0.006	37	-0.487	0.096	-5.08	<.001
3	0.349	0.115	3.05	0.002	38	-0.002	0.086	-0.03	0.978
4	0.112	0.085	1.32	0.188	40	0.148	0.152	0.97	0.333
5	-0.200	0.083	-2.41	0.016	43	-0.471	0.095	-4.94	<.001
6	-0.619	0.107	-5.77	<.001	44	0.204	0.084	2.42	0.016
7	0.098	0.105	0.94	0.349	45	0.152	0.099	1.54	0.124
8	-0.235	0.074	-3.20	0.001	46	0.044	0.102	0.43	0.665
10	-0.302	0.068	-4.45	<.001	47	0.018	0.094	0.19	0.850
11	0.327	0.095	3.45	0.001	48	-0.255	0.070	-3.63	0.000
12	-0.199	0.078	-2.54	0.011	49	-0.136	0.102	-1.33	0.185
13	-0.147	0.069	-2.11	0.035	50	-0.182	0.077	-2.35	0.019
14	0.079	0.107	0.74	0.459	51	-0.269	0.065	-4.16	<.001
15	-0.356	0.087	-4.11	<.001	52	0.017	0.113	0.15	0.879
16	0.105	0.095	1.1	0.272	53	0.209	0.107	1.96	0.050
17	0.187	0.109	1.72	0.086	55	0.187	0.102	1.82	0.069
18	0.122	0.082	1.49	0.137	56	0.335	0.131	2.55	0.011
19	0.055	0.097	0.57	0.569	57	0.163	0.126	1.30	0.195
20	-0.186	0.073	-2.55	0.011	58	0.051	0.107	0.47	0.637
22	0.069	0.084	0.82	0.411	60	0.048	0.131	0.37	0.712
23	-0.136	0.086	-1.58	0.113	61	-0.150	0.193	-0.78	0.436
24	-0.187	0.093	-2.02	0.044	64	0.330	0.075	4.40	<.001
25	0.270	0.103	2.62	0.009	65	0.226	0.078	2.88	0.004
26	0.193	0.123	1.58	0.115	66	0.501	0.072	6.93	<.001
27	-0.176	0.075	-2.35	0.019	70	0.054	0.074	0.73	0.465
28	0.160	0.110	1.46	0.146	71	0.100	0.083	1.20	0.230
29	-0.077	0.082	-0.94	0.348	74	-0.089	0.196	-0.45	0.651
30	-0.090	0.104	-0.87	0.387	75	-0.072	0.077	-0.93	0.353
31	0.359	0.096	3.75	0.000	77	0.026	0.076	0.34	0.733
32	-0.033	0.113	-0.29	0.772	78	0.164	0.193	0.85	0.394
33	-0.404	0.111	-3.65	<.001	109	-0.102	0.219	-0.46	0.642
34	0.002	0.073	0.03	0.974	110	-0.193	0.134	-1.45	0.148
35	-0.204	0.092	-2.21	0.027					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

Table 2.26b Intercept estimates for institutional assessment satisfaction effects

Institute	Estimate	Std err	t	Pr > |t|	Institute	Estimate	Std err	t	Pr > |t|
1	0.127	0.075	1.69	0.091	36	0.021	0.085	0.25	0.806
2	0.099	0.092	1.07	0.283	37	-0.365	0.093	-3.90	<.001
3	0.257	0.112	2.30	0.021	38	-0.038	0.084	-0.45	0.649
4	0.135	0.084	1.61	0.107	40	0.214	0.143	1.49	0.135
5	-0.277	0.082	-3.39	0.001	43	-0.157	0.089	-1.76	0.079
6	0.203	0.101	2.02	0.043	44	0.206	0.083	2.50	0.013
7	-0.081	0.103	-0.79	0.430	45	0.197	0.096	2.05	0.040
8	-0.283	0.072	-3.93	<.001	46	0.064	0.099	0.65	0.518
10	-0.382	0.066	-5.76	<.001	47	0.153	0.091	1.67	0.094
11	0.198	0.093	2.12	0.034	48	-0.130	0.068	-1.92	0.055
12	-0.171	0.077	-2.22	0.026	49	-0.110	0.099	-1.11	0.267
13	-0.172	0.068	-2.54	0.011	50	-0.025	0.074	-0.34	0.733
14	-0.116	0.102	-1.14	0.255	51	-0.165	0.062	-2.66	0.008
15	-0.342	0.084	-4.08	<.001	52	0.135	0.109	1.24	0.214
16	-0.016	0.093	-0.17	0.862	53	0.165	0.104	1.59	0.111
17	0.110	0.105	1.05	0.295	55	0.380	0.099	3.84	<.001
18	-0.055	0.080	-0.68	0.494	56	0.382	0.126	3.04	0.002
19	-0.099	0.095	-1.04	0.296	57	0.099	0.120	0.83	0.408
20	-0.228	0.071	-3.19	0.001	58	0.158	0.104	1.53	0.127
22	0.005	0.082	0.06	0.948	60	0.127	0.126	1.01	0.312
23	-0.125	0.083	-1.50	0.133	61	0.017	0.175	0.10	0.922
24	-0.336	0.090	-3.72	<.001	64	0.251	0.074	3.41	0.001
25	0.164	0.100	1.64	0.101	65	0.061	0.077	0.78	0.433
26	0.061	0.117	0.52	0.606	66	0.385	0.071	5.46	<.001
27	-0.244	0.072	-3.38	0.001	70	0.006	0.072	0.08	0.933
28	-0.024	0.106	-0.23	0.821	71	0.005	0.080	0.06	0.952
29	-0.051	0.079	-0.64	0.525	74	0.125	0.176	0.71	0.478
30	-0.270	0.102	-2.64	0.008	75	0.047	0.073	0.64	0.520
31	0.255	0.094	2.72	0.007	77	-0.062	0.073	-0.84	0.401
32	0.004	0.109	0.03	0.973	78	0.105	0.176	0.59	0.552
33	-0.274	0.107	-2.57	0.010	109	-0.287	0.194	-1.48	0.139
34	0.045	0.072	0.62	0.534	110	-0.058	0.128	-0.45	0.651
35	-0.024	0.089	-0.26	0.791					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

Table 2.26c Intercept estimates for institutional general learning satisfaction effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	0.158	0.075	2.10	0.036	36	-0.088	0.086	-1.03	0.305
2	0.207	0.092	2.24	0.025	37	-0.246	0.094	-2.62	0.009
3	0.294	0.112	2.63	0.009	38	-0.033	0.084	-0.39	0.698
4	0.198	0.083	2.37	0.018	40	0.266	0.144	1.85	0.065
5	-0.173	0.082	-2.12	0.034	43	-0.319	0.091	-3.52	<.001
6	-0.049	0.101	-0.49	0.627	44	0.178	0.083	2.13	0.033
7	0.156	0.103	1.53	0.127	45	0.161	0.096	1.67	0.096
8	0.220	0.072	3.04	0.002	46	0.057	0.100	0.57	0.572
10	-0.085	0.067	-1.28	0.201	47	0.004	0.092	0.04	0.966
11	0.243	0.093	2.61	0.009	48	0.037	0.068	0.54	0.587
12	0.019	0.077	0.25	0.801	49	0.094	0.098	0.95	0.340
13	-0.066	0.068	-0.98	0.329	50	-0.006	0.074	-0.08	0.939
14	-0.018	0.103	-0.18	0.860	51	-0.153	0.063	-2.45	0.014
15	-0.365	0.085	-4.31	<.001	52	0.135	0.110	1.22	0.223
16	-0.270	0.094	-2.87	0.004	53	-0.018	0.105	-0.18	0.860
17	-0.154	0.107	-1.45	0.148	55	0.213	0.100	2.13	0.033
18	-0.089	0.080	-1.11	0.265	56	0.091	0.127	0.72	0.472
19	-0.093	0.095	-0.97	0.331	57	0.054	0.121	0.45	0.653
20	-0.155	0.071	-2.17	0.030	58	0.095	0.105	0.91	0.362
22	0.104	0.083	1.24	0.214	60	0.036	0.125	0.29	0.773
23	-0.172	0.084	-2.04	0.042	61	0.021	0.174	0.12	0.904
24	-0.115	0.091	-1.26	0.207	64	0.108	0.074	1.46	0.145
25	0.199	0.101	1.97	0.049	65	-0.051	0.077	-0.66	0.508
26	0.146	0.118	1.24	0.214	66	0.219	0.071	3.07	0.002
27	-0.526	0.073	-7.18	<.001	70	-0.549	0.073	-7.55	<.001
28	0.044	0.107	0.41	0.681	71	0.098	0.080	1.21	0.225
29	0.036	0.080	0.45	0.651	74	0.260	0.175	1.49	0.137
30	0.041	0.102	0.40	0.686	75	0.103	0.074	1.39	0.164
31	-0.027	0.094	-0.29	0.774	77	-0.407	0.074	-5.49	<.001
32	0.102	0.110	0.93	0.352	78	0.232	0.175	1.33	0.185
33	0.007	0.107	0.06	0.949	109	0.005	0.192	0.03	0.978
34	-0.192	0.072	-2.67	0.008	110	-0.066	0.129	-0.52	0.606
35	-0.157	0.090	-1.74	0.081					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

As the model explaining teaching satisfaction was found to exhibit selection bias, and we subsequently adjusted for this by including the IMR in the final teaching satisfaction model, we were interested in what the nature of the selection bias was. Table 2.27 reports the estimates for the adjusted and unadjusted outcomes. It is evident that the only variable estimate that differed significantly between the adjusted and unadjusted model was 'age'. The IMR covariate was significant and positive, meaning that there was a

positive correlation between the error terms of the satisfaction and selection equations. This meant that there were unobserved factors that increased the response likelihood in respect to higher satisfaction with teaching. It can also be seen that the coefficient for age was much larger in the adjusted model. This suggested that response patterns were downward biasing satisfaction with increasing age.

Table 2.27 Teaching satisfaction estimates: adjusted v non-adjusted for selection bias

Effect	Heckman		No Heckman		Difference	t	Pr > t
	Estimate	Std error	Estimate	Std error	Estimate		
Intercept	3.383	0.246	4.304	0.205	0.921	2.88	0.004
Inverse Mills Ratio	0.399	0.090	.	.	.	.	.
Age	0.013	0.001	0.001	0.000	-0.011	-9.37	<.001
Sex							
Female	0.044	0.029	-0.015	0.026	-0.059	-1.50	0.134
Male	0.000	.	0.000	.	0.000	.	.
Group							
Graduate	0.850	0.039	0.804	0.038	-0.046	-0.84	0.401
Module Completer	0.000	.	0.000	.	0.000	.	.
Field of education							
Natural & Phys Sciences	0.230	0.253	0.196	0.254	-0.033	-0.09	0.928
Information Technology	-0.667	0.220	-0.713	0.220	-0.047	-0.15	0.881
Engineering & Related	-0.001	0.211	0.016	0.211	0.017	0.06	0.952
Architecture & Building	-0.171	0.214	-0.167	0.215	0.004	0.01	0.992
Agriculture & Related	0.362	0.215	0.402	0.215	0.040	0.13	0.897
Health	0.472	0.214	0.502	0.214	0.030	0.10	0.920
Education	0.217	0.214	0.249	0.215	0.032	0.11	0.912
Management & Commerce	0.092	0.211	0.096	0.212	0.004	0.01	0.992
Society & Culture	0.355	0.212	0.390	0.213	0.035	0.12	0.904
Creative Arts	0.154	0.216	0.107	0.217	-0.047	-0.15	0.881
Food, Hosp. & Pers. Serv.	0.299	0.213	0.311	0.213	0.011	0.04	0.968
Mixed Field Programs	-0.081	0.212	-0.040	0.213	0.042	0.14	0.889
Subject only enrolment	0.000	.	0.000	.	0.000	.	.
Qualification level							
Diploma & above	-1.214	0.074	-1.323	0.072	-0.109	-1.06	0.289
Certificate IV	-1.063	0.071	-1.137	0.070	-0.074	-0.74	0.459
Certificate III	-0.991	0.068	-1.072	0.067	-0.081	-0.85	0.395
Certificate II	-0.702	0.069	-0.770	0.069	-0.068	-0.7	0.484
Certificate I	-0.836	0.087	-0.821	0.086	0.015	0.12	0.904
Other	-0.470	0.067	-0.438	0.067	0.032	0.34	0.734
Statement of attainment	0.000	.	0.000	.	0.000	.	.

Note: Shaded predictors are significant.

2.1.18 Student overall satisfaction

In the Student Outcome Survey, students were asked to indicate their overall satisfaction in addition to the 19 questions dealing with the various sub-aspects of satisfaction. This overall satisfaction question conformed to a 1 (strong dissatisfaction) to 5 (strong satisfaction) Likert-style scale. Again we estimated two separate models predicting overall satisfaction using the institutional identifier as a random effect and our standard covariates plus the IMR in one of the models as fixed effects. Table 2.28 displays the type 3 fixed effects in both models. In the unadjusted model all covariates except for age were significant, while in the in the model which was adjusted for selection bias sex was non-significant. The very high value for the F-test for the 'Graduate/module completer' variable denoted that there were substantial differences between graduates and module completers with respect to overall satisfaction. The selection model contained the significant IMR, which pointed to the existence of selection bias. We therefore used the latter model to calculate institutional effects

Table 2.28 Type 3 effects on overall satisfaction

Variable	Basic			Basic + Mills	
	DF	F	Pr > F	F	Pr > F
Field of education (2-digit)	11	38.6	<.001	34.5	<.001
Age	1	0.1	0.759	5.2	0.022
Sex	1	15.0	0.003	2.6	0.113
Qualification level	6	90.9	<.001	68.0	<.001
Graduate/module completer	1	1002.3	<.001	1011.5	<.001
Inverse Mills ratio	1			18.9	<.001

The covariance parameter estimates (table 2.29) denoted statistically significant effects of TAFE institutes on overall satisfaction, after accounting for covariates.

Table 2.29 Covariance parameter estimates – overall satisfaction

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	0.003	0.001	4.38	<.001
Residual		0.765	0.004	176.08	<.001

To quantify the impact of institutional differences on overall satisfaction we ran an intercept-only model (table 2.30) and calculated the intra-class correlation using formula (7). The resulting ICC was 0.005, indicating that 0.5% of the variability in overall student satisfaction was explainable by grouping students with respect to the institutions they attended in addition to the variance between individual students.

Table 2.30 Covariance parameter estimates – intercept only

Parameter	Subject	Estimate	Std error	t	Pr > t
Intercept	Institute	0.004	0.001	4.53	<.001
Residual		0.788	0.004	176.16	<.001
Intraclass correlation			0.005		

While most of the differences in student satisfaction could be expected to stem from differences between individuals, we did find that a significant portion was due to institutional differences. However, it must be pointed out that the institutional differences in all four satisfaction models were less pronounced than in the employment and salary models evaluated earlier in this study. Estimates of individual institutes' intercepts can be found in table 2.31.

Table 2.31 Intercept estimates for institutional overall satisfaction effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	0.031	0.023	1.36	0.173	36	-0.020	0.026	-0.78	0.435
2	0.034	0.028	1.23	0.219	37	-0.133	0.028	-4.73	<.001
3	0.071	0.033	2.15	0.031	38	0.003	0.025	0.1	0.919
4	0.035	0.025	1.39	0.163	40	0.026	0.042	0.62	0.534
5	-0.057	0.024	-2.35	0.019	43	-0.090	0.027	-3.36	0.001
6	-0.028	0.031	-0.92	0.356	44	0.048	0.025	1.95	0.052
7	0.004	0.030	0.12	0.908	45	0.055	0.029	1.89	0.059
8	-0.020	0.022	-0.90	0.367	46	0.008	0.030	0.28	0.783
10	-0.078	0.020	-3.87	0.000	47	-0.005	0.028	-0.17	0.861
11	0.069	0.028	2.49	0.013	48	-0.044	0.021	-2.11	0.035
12	-0.081	0.023	-3.5	0.001	49	0.007	0.030	0.25	0.806
13	-0.045	0.021	-2.21	0.027	50	-0.043	0.023	-1.87	0.061
14	-0.012	0.031	-0.38	0.706	51	-0.044	0.019	-2.32	0.020
15	-0.087	0.025	-3.41	0.001	52	0.032	0.033	0.98	0.328
16	0.012	0.028	0.44	0.662	53	0.031	0.031	1.02	0.310
17	0.023	0.031	0.74	0.460	55	0.061	0.030	2.03	0.042
18	0.032	0.024	1.33	0.184	56	0.073	0.037	1.97	0.049
19	-0.032	0.028	-1.14	0.255	57	0.068	0.036	1.9	0.058
20	-0.061	0.022	-2.85	0.004	58	0.039	0.031	1.24	0.214
22	-0.016	0.025	-0.65	0.515	60	-0.035	0.037	-0.94	0.348
23	-0.015	0.025	-0.61	0.544	61	0.006	0.049	0.12	0.906
24	-0.063	0.027	-2.32	0.021	64	0.098	0.022	4.42	<.001
25	0.090	0.030	2.97	0.003	65	0.030	0.023	1.3	0.195
26	0.056	0.035	1.6	0.109	66	0.109	0.021	5.12	<.001
27	-0.099	0.022	-4.53	<.001	70	0.004	0.022	0.20	0.838
28	-0.004	0.032	-0.12	0.906	71	-0.019	0.025	-0.77	0.442
29	-0.021	0.024	-0.86	0.389	74	0.002	0.050	0.04	0.971
30	-0.021	0.030	-0.71	0.480	75	-0.022	0.023	-0.97	0.331
31	0.069	0.028	2.45	0.014	77	-0.005	0.022	-0.24	0.811
32	0.011	0.033	0.32	0.746	78	0.012	0.049	0.24	0.811
33	-0.018	0.032	-0.57	0.571	109	0.007	0.054	0.13	0.899
34	-0.006	0.021	-0.28	0.783	110	-0.010	0.038	-0.27	0.788
35	-0.020	0.027	-0.74	0.460					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

A graphical representation of overall satisfaction with training by institution can be seen in figure 2.5.

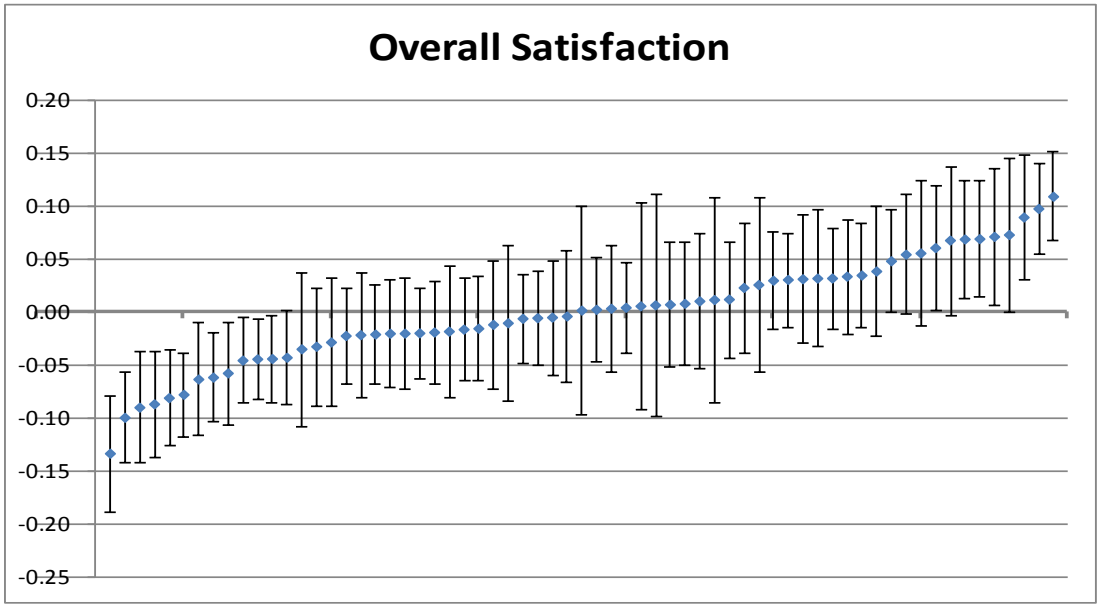

Figure 2.5 Intercept estimates for institutional overall satisfaction effects

A comparison of the two overall satisfaction models in respect to the impact of the detected selection bias (table 2.32) revealed that selection bias was confined to the age variable, which was statistically different when both models were compared. In fact, in the unadjusted model, the age variable itself had no notable predictive value for overall student satisfaction (table 2.28). We can therefore argue that considering an overall satisfaction model without adjusting for selection bias may lead to model specification errors if student age is included as a prediction variable. This underlines the necessity to address non-response bias in this type of survey.

Table 2.32 Overall satisfaction estimates: adjusted v non-adjusted for selection bias

Effect	Heckman		No Heckman		Difference	t	Pr > t
	Estimate	Std error	Estimate	Std error	Estimate		
Intercept	4.136	0.0758	4.307	0.0628	0.171	1.73	0.084
Inverse Mills ratio	0.120	0.028	.	.	.	.	.
Age	0.001	0.000	0.000	0.000	-0.001	-2.27	0.023
Sex							
Female	-0.015	0.009	-0.032	0.008	-0.017	-1.40	0.162
Male	0.000	.	0.000	.	0.000	.	.
Group							
Graduate	0.389	0.012	0.379	0.012	-0.009	-0.55	0.582
Module Completers	0.000	.	0.000	.	0.000	.	.
Field of education							
Natural & Phys Sciences	0.021	0.079	0.017	0.079	-0.004	-0.03	0.976
Information Technology	-0.378	0.068	-0.389	0.068	-0.010	-0.11	0.912
Engineering & Related	-0.067	0.065	-0.058	0.065	0.009	0.10	0.92
Architecture & Building	-0.088	0.066	-0.076	0.066	0.012	0.13	0.897
Agriculture & Related	0.005	0.066	0.014	0.067	0.009	0.10	0.920
Health	0.037	0.066	0.043	0.066	0.005	0.05	0.960
Education	-0.043	0.066	-0.046	0.066	-0.003	-0.03	0.976
Management & Commerce	-0.021	0.065	-0.018	0.066	0.004	0.04	0.968
Society & Culture	0.051	0.066	0.058	0.066	0.007	0.08	0.936
Creative Arts	-0.033	0.067	-0.033	0.067	-0.001	-0.01	0.992
Food, Hosp. & Pers. Serv.	0.033	0.066	0.050	0.066	0.017	0.18	0.857
Mixed Field Programs	-0.033	0.066	-0.019	0.066	0.014	0.15	0.881
Subject only enrolment	0.000	.	0.000	.	0.000	.	.
Qualification level							
Diploma & above	-0.413	0.023	-0.437	0.023	-0.025	-0.76	0.447
Certificate IV	-0.401	0.022	-0.419	0.022	-0.018	-0.58	0.562
Certificate III	-0.356	0.021	-0.365	0.021	-0.010	-0.32	0.749
Certificate II	-0.272	0.022	-0.274	0.022	-0.002	-0.07	0.944
Certificate I	-0.288	0.027	-0.275	0.027	0.014	0.36	0.719
Other	-0.145	0.021	-0.137	0.021	0.008	0.27	0.787
Statement of attainment	0.000	.	0.000	.	0.000	.	.

Note: Shaded predictors are significant.

2.1.19 Student perception of achievement and willingness to recommend their institution

The last two outcomes we considered in this study were students' perceptions of achievement and the willingness to recommend their institution of training. The SOS asked sampled students: 'Did the training help you to achieve your main reason [for doing the training]?'. The possible affirmative and negative responses to this question include 'yes', 'no' and 'partly'. For the purpose of our analysis we recoded this variable into two categories, comprising

'yes/partly', and 'no'. This yielded a dichotomous variable which we then analysed via a logistic hierarchical model.

The other outcome to be investigated was the willingness of students to recommend their institution. The SOS asked 'Would you recommend the training provider where you undertook the training to others?'. This was an outright dichotomous variable with 'yes' or 'no' as possible answers.

The methodology we employed here somewhat resembled the approach from our employment model. This meant that a hierarchical logistic regression model was used. The fixed effect variables entered consisted only of the standard covariates of field of education, age, sex, qualification level and group, as well as the IMR to test and adjust for selection bias. The only random effect was the institution itself. Type 3 effects on both outcomes can be found in table 2.33. The models for both outcomes displayed a significant IMR, pointing to selection bias. With the exception of the sex variable in the 'achieved' model, all independent fixed effects were significant. As in previous models, the high F-test value for the module graduate/completer indicator showed that these two student categories differed strongly in their response to both questions.

Table 2.33 Type 3 effects on 'achieved' and 'recommend' outcomes

		Achieved				**Recommend**			
Variable		Basic		Basic + Mills		Basic		Basic + Mills	
	DF	*F*	*Pr > F*	*F*	*Pr > F*	*F*	*Pr > F*	*F*	*Pr > F*
Field of education	11	31.8	<.001	27.4	<.001	11.1	<.001	9.2	<.001
Age	1	1.1	0.286	5.3	0.022	2.2	0.139	27.6	<.001
Sex	1	4.3	0.043	0.3	0.589	54.9	<.001	28.9	<.001
Qualification level	6	105.6	<.001	93.3	<.001	56.2	<.001	41.3	<.001
Graduate/MC	1	1650.4	<.001	1550.9	<.001	509.6	<.001	507.9	<.001
Inverse Mills ratio				7.2	0.007			9.5	0.002

Covariance parameter estimates for both models including the IMR were significant (table 2.34), indicating the existence of institutional effects on these two outcomes.

Table 2.34 Covariance parameters for 'achieved' and 'recommend'

		Achieved		Recommend	
Parameter	Subject	Estimate	Pr > t	Estimate	Pr > t
Intercept	Institute	0.063	<.0001	0.027	0.001

To estimate the amount of variance in the 'achievement' and 'recommend' variable that can be explained by grouping students within the institutions they have attended, the intraclass coefficients were calculated by re-running both models as intercept-only models. The results (table 2.35) indicated 2.4% (achieved) and 1% (recommend) of variability can be explained by difference between institutions.

Table 2.35 Covariance parameter estimates – intercept only

		Achieved		Recommend	
		Estimate	Pr > t	Estimate	Pr > t
Intercept	*Institute*	*0.081*	*<.0001*	*0.034*	*0.001*
Intraclass correlation		0.024		0.010	

We included the results of institutional intercepts for 'achieved' and 'recommend' in tables 2.36 and 2.37.

Table 2.36 Intercept estimates for 'achieved' effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	-0.347	0.097	-3.57	<.001	36	-0.296	0.114	-2.61	0.009
2	-0.087	0.125	-0.70	0.487	37	-0.588	0.118	-5.00	<.001
3	0.033	0.151	0.22	0.827	38	-0.327	0.108	-3.02	0.003
4	-0.164	0.108	-1.53	0.127	40	-0.090	0.188	-0.48	0.632
5	-0.140	0.109	-1.28	0.201	43	-0.057	0.125	-0.46	0.645
6	-0.078	0.133	-0.59	0.556	44	0.251	0.126	2.00	0.046
7	0.081	0.139	0.58	0.563	45	-0.015	0.132	-0.11	0.910
8	-0.289	0.094	-3.06	0.002	46	-0.272	0.129	-2.11	0.035
10	-0.263	0.086	-3.05	0.002	47	-0.059	0.135	-0.43	0.664
11	0.115	0.129	0.89	0.373	48	-0.002	0.100	-0.02	0.981
12	-0.225	0.099	-2.26	0.024	49	-0.059	0.137	-0.43	0.669
13	-0.341	0.088	-3.86	0.000	50	-0.082	0.109	-0.75	0.451
14	0.035	0.139	0.25	0.801	51	-0.132	0.086	-1.53	0.125
15	0.088	0.120	0.73	0.465	52	0.244	0.164	1.49	0.137
16	0.118	0.133	0.89	0.373	53	0.078	0.148	0.53	0.597
17	0.319	0.161	1.98	0.048	55	0.104	0.140	0.74	0.458
18	0.146	0.118	1.23	0.218	56	0.207	0.176	1.18	0.239
19	0.261	0.146	1.78	0.075	57	0.315	0.180	1.75	0.081
20	-0.078	0.102	-0.76	0.445	58	0.166	0.155	1.07	0.283
22	0.004	0.117	0.03	0.976	60	-0.070	0.174	-0.40	0.689
23	-0.288	0.111	-2.60	0.009	61	0.161	0.223	0.72	0.470
24	-0.116	0.129	-0.89	0.372	64	-0.275	0.100	-2.75	0.006
25	0.367	0.159	2.31	0.021	65	-0.313	0.099	-3.16	0.002
26	-0.118	0.156	-0.75	0.451	66	-0.114	0.099	-1.15	0.249
27	0.013	0.102	0.12	0.903	70	0.311	0.115	2.71	0.007
28	-0.071	0.142	-0.50	0.619	71	-0.232	0.106	-2.20	0.028
29	-0.079	0.112	-0.70	0.482	74	0.046	0.230	0.20	0.840
30	0.261	0.147	1.78	0.076	75	0.415	0.120	3.46	0.001
31	0.367	0.146	2.53	0.012	77	0.159	0.106	1.51	0.132
32	0.267	0.157	1.70	0.090	78	0.002	0.244	0.01	0.993
33	0.139	0.150	0.93	0.352	109	0.102	0.242	0.42	0.673
34	-0.137	0.096	-1.42	0.156	110	0.179	0.197	0.90	0.366
35	0.182	0.132	1.38	0.167					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

Table 2.37 Intercept estimates for 'recommend' effects

Institute	Estimate	Std error	t	Pr > |t|	Institute	Estimate	Std error	t	Pr > |t|
1	0.108	0.090	1.20	0.230	36	0.012	0.098	0.13	0.899
2	0.103	0.106	0.97	0.331	37	-0.258	0.102	-2.54	0.011
3	0.128	0.122	1.05	0.292	38	-0.250	0.093	-2.7	0.007
4	0.145	0.099	1.47	0.142	40	-0.114	0.139	-0.82	0.413
5	-0.028	0.092	-0.31	0.759	43	-0.028	0.100	-0.28	0.779
6	0.101	0.112	0.91	0.365	44	0.093	0.098	0.95	0.344
7	0.132	0.115	1.15	0.250	45	-0.085	0.107	-0.80	0.425
8	-0.190	0.081	-2.35	0.019	46	-0.046	0.110	-0.42	0.677
10	-0.115	0.075	-1.54	0.123	47	0.024	0.107	0.22	0.825
11	0.254	0.115	2.21	0.027	48	-0.071	0.080	-0.88	0.378
12	-0.184	0.084	-2.18	0.029	49	0.153	0.115	1.33	0.183
13	-0.178	0.078	-2.28	0.023	50	-0.114	0.088	-1.29	0.196
14	-0.029	0.113	-0.26	0.794	51	0.034	0.073	0.47	0.641
15	-0.214	0.093	-2.29	0.022	52	0.056	0.123	0.46	0.648
16	0.053	0.107	0.50	0.619	53	0.108	0.118	0.91	0.361
17	0.036	0.118	0.31	0.759	55	0.117	0.113	1.03	0.302
18	0.129	0.097	1.33	0.184	56	0.084	0.133	0.63	0.527
19	-0.121	0.106	-1.14	0.254	57	0.092	0.131	0.71	0.481
20	-0.146	0.083	-1.76	0.078	58	0.044	0.117	0.38	0.707
22	-0.070	0.096	-0.73	0.466	60	-0.102	0.129	-0.79	0.431
23	-0.041	0.096	-0.43	0.670	61	0.028	0.152	0.19	0.853
24	-0.107	0.099	-1.08	0.282	64	0.170	0.090	1.88	0.060
25	0.235	0.123	1.92	0.055	65	0.152	0.092	1.65	0.099
26	0.057	0.125	0.46	0.647	66	0.187	0.089	2.09	0.036
27	-0.286	0.083	-3.45	0.001	70	0.113	0.092	1.23	0.219
28	-0.040	0.114	-0.35	0.724	71	-0.030	0.092	-0.33	0.740
29	-0.135	0.091	-1.49	0.137	74	-0.072	0.157	-0.46	0.648
30	-0.003	0.113	-0.03	0.976	75	-0.204	0.088	-2.31	0.021
31	0.177	0.114	1.55	0.121	77	0.022	0.085	0.25	0.799
32	0.007	0.118	0.06	0.951	78	-0.004	0.160	-0.02	0.980
33	-0.118	0.114	-1.04	0.299	109	0.072	0.160	0.45	0.651
34	0.008	0.084	0.09	0.925	110	0.020	0.135	0.15	0.883
35	-0.020	0.103	-0.20	0.843					

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

The core information in the above tables, for example, estimates and confidence interval for the 'achieved' and 'recommend' outcomes, is summarised in the following two figures (2.6 and 2.7)

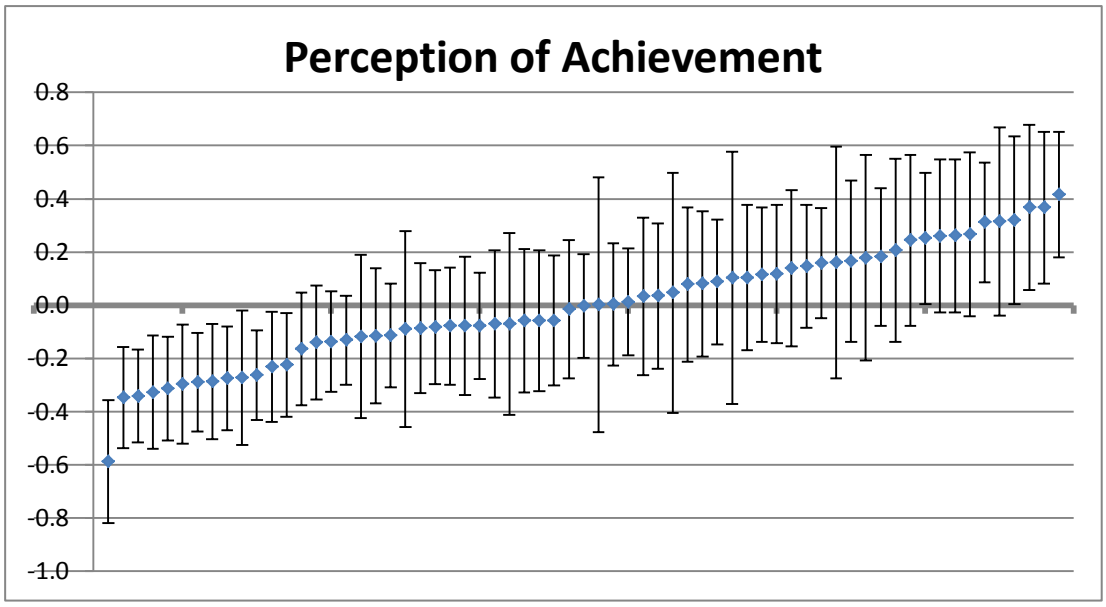

Figure 2.6 Intercept estimates for 'achieved' effects

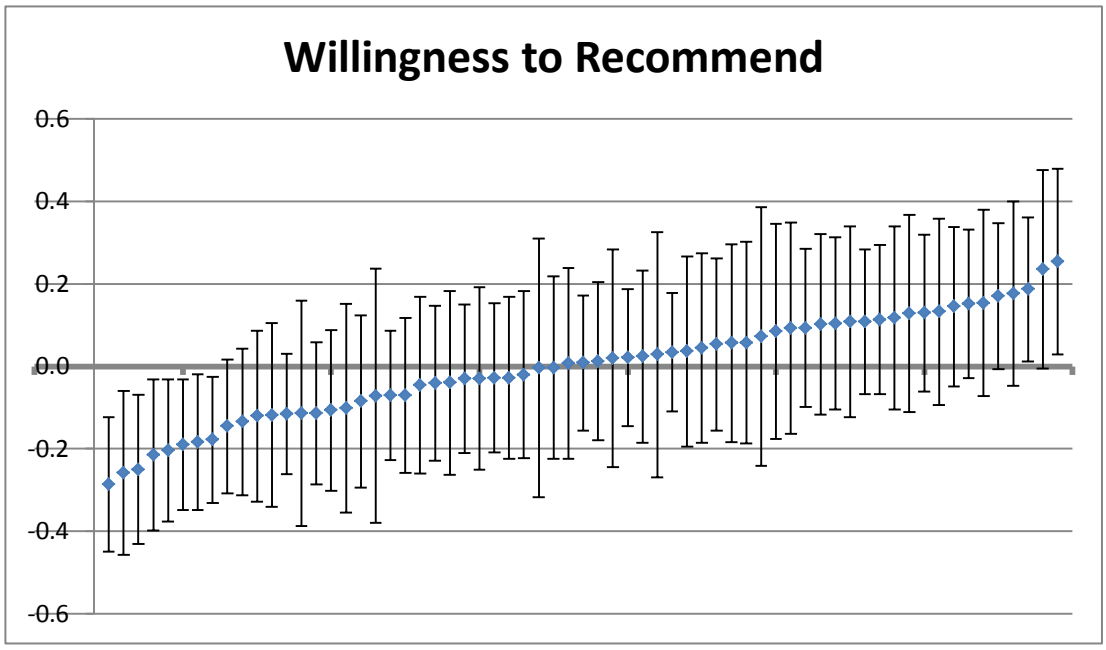

Figure 2.7 Intercept estimates for 'recommend' effects

Our initial test of the significance of the IMR denoted the existence of selection bias in both the achievement and the recommended models. A comparison between the adjusted and unadjusted model (tables 2.38 and 2.39) showed that only the age variable was moderately significantly different. In both adjusted models the coefficient for age was much larger than in the unadjusted model indicating that selection effects lead to lower estimates of perceptions of achievement and willingness to recommend the institution.

Table 2.38 'Achieved' estimates: adjusted v non-adjusted for selection bias

Effect	Heckman		No Heckman		Diff	t	Pr > |t|
	Estimate	Std error	Estimate	Std error			
Intercept	2.021	0.251	2.592	0.103	0.571	2.11	0.035
Inverse Mills ratio	0.369	0.139	.	.	.	.	.
Age	0.004	0.002	0.001	0.001	-0.003	-1.78	0.075
Sex							
Female	-0.024	0.043	-0.078	0.038	-0.054	-0.95	0.342
Male	0.000	.	0.000	.	0.000	.	.
Group							
Graduate	1.629	0.041	1.598	0.039	-0.031	-0.54	0.589
Module Completer	0.000	.	0.000	.	0.000	.	.
Field of education							
Natural & Phys Sciences	-0.881	0.176	-0.894	0.174	-0.013	-0.05	0.960
Information Technology	-0.665	0.099	-0.697	0.094	-0.032	-0.23	0.818
Engineering & Related	0.222	0.078	0.261	0.077	0.039	0.36	0.719
Architecture & Building	0.155	0.095	0.196	0.094	0.041	0.31	0.757
Agriculture & Related	0.528	0.108	0.605	0.109	0.077	0.5	0.617
Health	0.297	0.088	0.320	0.087	0.023	0.19	0.849
Education	-0.038	0.092	-0.032	0.090	0.006	0.05	0.960
Management & Commerce	-0.407	0.068	-0.395	0.066	0.012	0.13	0.897
Society & Culture	0.123	0.077	0.155	0.076	0.032	0.29	0.772
Creative Arts	-0.193	0.095	-0.189	0.093	0.004	0.03	0.976
Food, Hosp. & Pers. Serv.	0.100	0.083	0.152	0.083	0.052	0.44	0.660
Mixed Field Programs	0.000	.	0.000	.	0.000	.	.
Qualification level							
Diploma & above	-1.154	0.100	-1.246	0.097	-0.093	-0.67	0.503
Certificate IV	-1.155	0.095	-1.231	0.093	-0.076	-0.57	0.569
Certificate III	-1.199	0.091	-1.244	0.090	-0.046	-0.35	0.726
Certificate II	-1.423	0.092	-1.449	0.092	-0.026	-0.2	0.841
Certificate I	-1.235	0.122	-1.195	0.121	0.040	0.24	0.810
Other	0.219	0.098	0.246	0.097	0.027	0.2	0.841
Statement of attainment	0.000	.	0.000	.	0.000	.	.

Table 2.39 'Recommend' estimates: adjusted v non-adjusted for selection bias

Effect	Heckman		No Heckman		Diff	t	Pr > |t|
	Estimate	Std error	Estimate	Std error			
Intercept	2.020	0.219	2.720	0.099	0.700	2.91	0.004
Inverse Mills ratio	0.369	0.119	.	.	.	.	.
Age	0.008	0.002	0.001	0.001	-0.007	-4.28	<.001
Sex							
Female	-0.213	0.040	-0.264	0.036	-0.051	-0.97	0.332
Male	0.000	.	0.000	.	0.000	.	.
Group							
Graduate	0.946	0.042	0.913	0.040	-0.033	-0.57	0.569
Module Completer	0.000	.	0.000	.	0.000	.	.
Field of education							
Natural & Phys Sciences	0.228	0.193	0.153	0.191	-0.075	-0.28	0.779
Information Technology	-0.396	0.092	-0.477	0.088	-0.081	-0.64	0.522
Engineering & Related	0.080	0.070	0.081	0.069	0.001	0.01	0.992
Architecture & Building	0.236	0.087	0.231	0.087	-0.006	-0.05	0.960
Agriculture & Related	0.286	0.092	0.307	0.091	0.021	0.16	0.873
Health	0.224	0.076	0.219	0.076	-0.006	-0.05	0.960
Education	0.067	0.081	0.054	0.079	-0.013	-0.11	0.912
Management & Commerce	0.241	0.065	0.222	0.063	-0.020	-0.22	0.826
Society & Culture	0.360	0.070	0.358	0.070	-0.002	-0.02	0.984
Creative Arts	0.143	0.087	0.093	0.086	-0.051	-0.41	0.682
Food, Hosp. & Pers. Serv.	0.283	0.078	0.286	0.078	0.002	0.02	0.984
Mixed Field Programs	0.000	.	0.000	.	0.000	.	.
Qualification level							
Diploma & above	-1.273	0.100	-1.326	0.096	-0.053	-0.38	0.704
Certificate IV	-1.157	0.096	-1.184	0.094	-0.027	-0.2	0.841
Certificate III	-1.069	0.093	-1.092	0.092	-0.023	-0.18	0.857
Certificate II	-0.872	0.096	-0.880	0.095	-0.008	-0.06	0.952
Certificate I	-0.946	0.122	-0.885	0.120	0.061	0.35	0.726
Other	-0.255	0.098	-0.178	0.097	0.077	0.56	0.575
Statement of attainment	0.000	.	0.000	.	0.000	.	.

2.1.20 Performance indicators compared

After creating eight distinct performance indicators, each adjusted for demographic and educational profiles of individual institution, we were interested in how these indicators compared, and whether there were patterns that could be discerned from looking at the indicators side by side. To do this, we used the parameter estimates for each institution, including the designation of above and below mean performance and joined them in table 2.40.

Table 2.40 Eight performance indicators compared side by side

Institute	Employed	Salary	Satis teach	Satis ass	Satis gen	Satis total	Achieved	Recom mend
1	-0.135	211	0.146	0.127	0.158	0.031	-0.347	0.108
2	-0.059	215	0.260	0.099	0.207	0.034	-0.087	0.103
3	-0.010	-1476	0.349	0.257	0.294	0.071	0.033	0.128
4	-0.275	-1965	0.112	0.135	0.198	0.035	-0.164	0.145
5	-0.114	-846	-0.200	-0.277	-0.173	-0.057	-0.140	-0.028
6	-0.119	-1555	-0.619	0.203	-0.049	-0.028	-0.078	0.101
7	0.032	-2188	0.098	-0.081	0.156	0.004	0.081	0.132
8	-0.404	45	-0.235	-0.283	0.220	-0.020	-0.289	-0.190
10	-0.182	-229	-0.302	-0.382	-0.085	-0.078	-0.263	-0.115
11	-0.129	1040	0.327	0.198	0.243	0.069	0.115	0.254
12	-0.096	796	-0.199	-0.171	0.019	-0.081	-0.225	-0.184
13	0.144	-1833	-0.147	-0.172	-0.066	-0.045	-0.341	-0.178
14	0.002	-2966	0.079	-0.116	-0.018	-0.012	0.035	-0.029
15	0.165	-2200	-0.356	-0.342	-0.365	-0.087	0.088	-0.214
16	0.086	-1571	0.105	-0.016	-0.270	0.012	0.118	0.053
17	-0.002	-3035	0.187	0.110	-0.154	0.023	0.319	0.036
18	-0.010	-2049	0.122	-0.055	-0.089	0.032	0.146	0.129
19	0.242	-1277	0.055	-0.099	-0.093	-0.032	0.261	-0.121
20	-0.082	-2713	-0.186	-0.228	-0.155	-0.061	-0.078	-0.146
22	0.014	-4489	0.069	0.005	0.104	-0.016	0.004	-0.070
23	-0.104	-1953	-0.136	-0.125	-0.172	-0.015	-0.288	-0.041
24	-0.166	-840	-0.187	-0.336	-0.115	-0.063	-0.116	-0.107
25	-0.006	-2210	0.270	0.164	0.199	0.090	0.367	0.235
26	-0.096	-2999	0.193	0.061	0.146	0.056	-0.118	0.057
27	0.045	3531	-0.176	-0.244	-0.526	-0.099	0.013	-0.286
28	0.141	1177	0.160	-0.024	0.044	-0.004	-0.071	-0.040
29	-0.195	275	-0.077	-0.051	0.036	-0.021	-0.079	-0.135
30	0.267	-1940	-0.090	-0.270	0.041	-0.021	0.261	-0.003
31	0.154	-1958	0.359	0.255	-0.027	0.069	0.367	0.177
32	0.096	-1024	-0.033	0.004	0.102	0.011	0.267	0.007
33	0.031	-1133	-0.404	-0.274	0.007	-0.018	0.139	-0.118
34	0.019	2245	0.002	0.045	-0.192	-0.006	-0.137	0.008
35	0.134	5599	-0.204	-0.024	-0.157	-0.020	0.182	-0.020
36	0.025	-665	0.089	0.021	-0.088	-0.020	-0.296	0.012
37	-0.244	-1711	-0.487	-0.365	-0.246	-0.133	-0.588	-0.258
38	-0.254	-1834	-0.002	-0.038	-0.033	0.003	-0.327	-0.250
40	-0.057	31	0.148	0.214	0.266	0.026	-0.090	-0.114
43	-0.032	-19	-0.471	-0.157	-0.319	-0.090	-0.057	-0.028
44	0.144	-758	0.204	0.206	0.178	0.048	0.251	0.093
45	-0.306	515	0.152	0.197	0.161	0.055	-0.015	-0.085
46	0.164	200	0.044	0.064	0.057	0.008	-0.272	-0.046
47	0.000	-448	0.018	0.153	0.004	-0.005	-0.059	0.024
48	-0.011	2514	-0.255	-0.130	0.037	-0.044	-0.002	-0.071
49	0.074	-490	-0.136	-0.110	0.094	0.007	-0.059	0.153
50	-0.226	4934	-0.182	-0.025	-0.006	-0.043	-0.082	-0.114
51	-0.056	1429	-0.269	-0.165	-0.153	-0.044	-0.132	0.034

Continued next page

Institute	Employed	Salary	Satis teach	Satis ass	Satis gen	Satis total	Achieved	Recom mend
52	-0.044	4117	0.017	0.135	0.135	0.032	0.244	0.056
53	0.005	1006	0.209	0.165	-0.018	0.031	0.078	0.108
55	0.130	1860	0.187	0.380	0.213	0.061	0.104	0.117
56	-0.012	7302	0.335	0.382	0.091	0.073	0.207	0.084
57	0.070	2194	0.163	0.099	0.054	0.068	0.315	0.092
58	0.204	9770	0.051	0.158	0.095	0.039	0.166	0.044
60	0.088	2979	0.048	0.127	0.036	-0.035	-0.070	-0.102
61	0.180	608	-0.150	0.017	0.021	0.006	0.161	0.028
64	0.100	-1494	0.330	0.251	0.108	0.098	-0.275	0.170
65	-0.010	-849	0.226	0.061	-0.051	0.030	-0.313	0.152
66	-0.037	-3362	0.501	0.385	0.219	0.109	-0.114	0.187
70	0.329	1379	0.054	0.006	-0.549	0.004	0.311	0.113
71	-0.230	-2168	0.100	0.005	0.098	-0.019	-0.232	-0.030
74	-0.040	831	-0.089	0.125	0.260	0.002	0.046	-0.072
75	0.117	2309	-0.072	0.047	0.103	-0.022	0.415	-0.204
77	0.189	1411	0.026	-0.062	-0.407	-0.005	0.159	0.022
78	-0.077	-3339	0.164	0.105	0.232	0.012	0.002	-0.004
109	-0.012	258	-0.102	-0.287	0.005	0.007	0.102	0.072
110	0.139	806	-0.193	-0.058	-0.066	-0.010	0.179	0.020

Notes: = Institutes performing significantly above the mean.
= Institutes performing significantly below the mean.

Visual inspection of this table shows that profiles above or below average performance were often associated with individual institutions. For instance, institutes 2, 3, 11, 44, 55, 56 and 66 scored mostly above average, whereas institutes 5, 10, 12, 13, 15, 20, 23, 24, 37, 38, 51 and 71 scored mostly below average. This indicated that it was indeed possible to identify institutions that consistently performed on either side of the spectrum.

It is of course possible to argue that some of the performance indicators reflect the same performance construct. Specifically the various satisfaction items could be suspected of being fairly correlated with each other. A cursory look at table 2.40 seems to confirm this suspicion. For instance, institutions 3, 5, 11 and 15 scored significantly above or below on all four satisfaction constructs. While it may be necessary to utilise all four satisfaction indicators in a detailed analysis of student satisfaction, in an overall investigation of performance indicators the presence of 50% of satisfaction related indicators

may be considered overbearing. We therefore wanted to know if some of the indicators measured statistically similar traits and if we could therefore reduce the number of indicators to a more manageable number that allowed clear allocation of identifiable concepts.

Looking at the correlation (table 2.41) of all indicators, it became clear that several indicators were highly correlated. While, for instance, 'salary' appears to be largely on its own, the satisfaction indicators all correlated highly with each other. The 'recommend' indicator was also highly related to three of the 'satisfaction' indicators. This, of course, made sense, as one could assume that students would often recommend their institution if they were satisfied with it. It was also noteworthy that the 'achieved' indicator was highly associated with the 'employed' indicator. This signified that students considered obtaining employment as one of their main goals when embarking on an education in the TAFE system.

Table 2.41 Correlations between individual performance indicators

	Employed	Salary	SatisT	SatisA	SatisG	SatisO	Achiev	Recomm
Employed	1							
Salary	0.17	1						
Teaching Satis	0.15	-0.07	1					
Assessment Satis	0.11	0.17	0.73	1				
Learning Satis	-0.27	-0.06	0.47	0.53	1			
Overall Satis	0.12	0.00	0.84	0.82	0.59	1		
Achieved	0.57	0.21	0.23	0.26	0.01	0.34	1	
Recommend	0.21	-0.05	0.60	0.61	0.33	0.77	0.31	1

Note: Shaded correlations are >0.5.

To further explore any underlying dimensionality underpinning our eight performance indicators we performed a principal component analysis with promax rotation of components. The eigenvalues of the correlation matrix (table 2.42) indicated two components with eigenvalues larger than one.

Table 2.42 Eigenvalues of the correlation matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	3.72	2.03	0.47	0.47
2	1.70	0.72	0.21	0.68
3	0.98	0.42	0.12	0.80
4	0.55	0.12	0.07	0.87
5	0.43	0.12	0.05	0.92
6	0.31	0.09	0.04	0.96
7	0.22	0.14	0.03	0.99
8	0.08	0.01	1.00	

As the third eigenvalue was fairly close to one, we also ran a parallel analysis to compare actual and simulated eigenvalues. The result appeared to confirm the existence of two dimensions (figure 2.8)

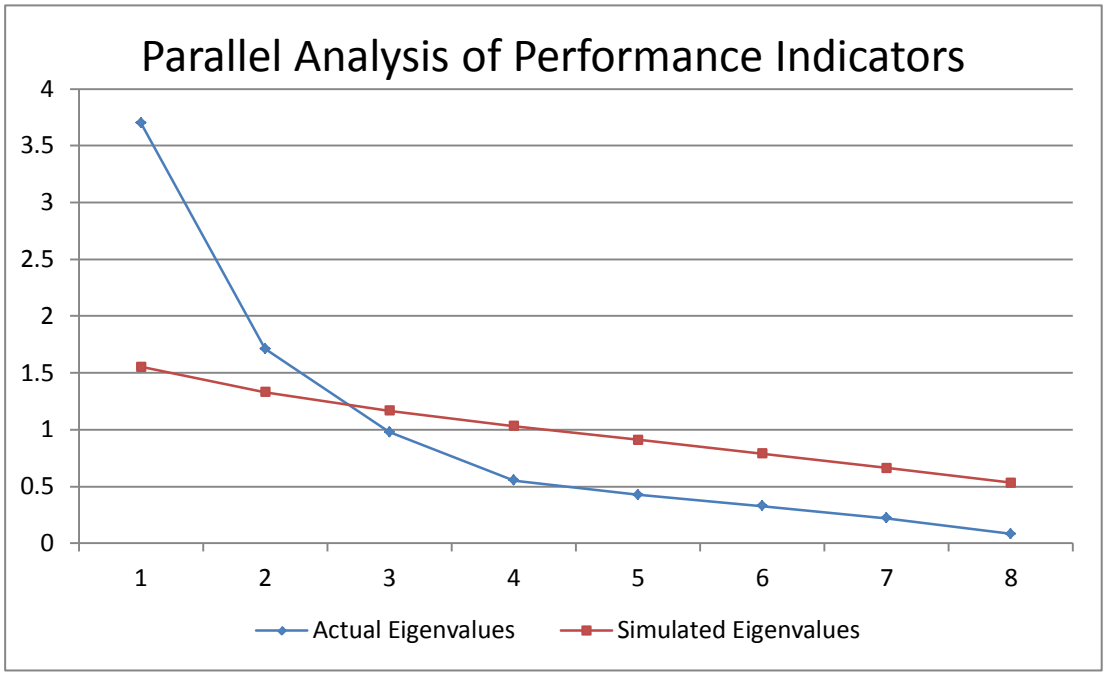

Figure 2.8 Parallel analysis of performance indicators

We subsequently extracted two components and considered the rotated factor pattern. In this pattern (table 2.43) it appeared that salary did not correlate well with either of the two components. All other indicators correlated strongly (>0.7) with one of the two components.

Table 2.43 Factor pattern (standardized regression coefficients) after Promax rotation (two retained factors)

Indicator	Factor 1	Factor 2
Employed after training	-0.033	0.871
Salary	-0.083	0.458
Teaching Satisfaction	0.872	0.012
Assessment Satisfaction	0.870	0.059
Learning Satisfaction	0.744	-0.435
Overall Satisfaction	0.959	0.043
Achieved	0.200	0.778
Recommend	0.764	0.158

Note: Shaded correlations are >0.5.

It seemed therefore reasonable to extract a third factor to account for the relative uniqueness of the salary indicator. The updated rotated factor pattern can be seen in table 2.44. All indicators correlated strongly with one of the three components. The shaded values indicate where strong relationships between a component and indicators existed.

Table 2.44 Factor pattern (standardized regression coefficients) after Promax rotation (three retained factors)

Indicator	Factor 1	Factor 2	Factor 3
Employed	-0.062	0.910	0.002
Salary	-0.039	0.098	0.967
Teaching Satisfaction	0.862	0.076	-0.116
Assessment Satisfaction	0.878	0.001	0.200
Learning Satisfaction	0.761	-0.472	0.082
Overall Satisfaction	0.952	0.075	-0.028
Achieved	0.182	0.762	0.143
Recommend	0.745	0.252	-0.180

Note: Shaded correlations are >0.5.

It became clear that our eight performance indicators could essentially be grouped into three coherent higher-order categories. These categories were

- Satisfaction — comprising the various satisfaction measures plus the 'willingness to recommend' variable. This seemed perfectly sensible, as one would expect a fairly large degree of coherence with different types of student satisfaction with their training. Additionally, the 'willingness to recommend' the institution could be seen as an indirect reflection of student satisfaction.

- Sense of achievement/employment outcomes — This dimension demonstrates that for students the sense of 'achievement' is intrinsically linked to success in securing a job after their training.
- Salary — This indicator stood completely on its own and, somewhat surprisingly, was not related to 'achievement'.

Of these three components, the satisfaction/willingness to recommend component could be interpreted as a 'subjectively' based performance indicator, whereas the sense of achievement/employment outcomes and salary components could be interpreted as more 'objectively' based indicators. The result of the principal component analysis of our performance indicators suggested that it is possible to capture 80% of the variance by reducing the number of indicators from eight to three (see table 2.42).

However, we submit that it would be more advantageous to retain the eight indicators that we have identified in this paper. The primary reason is that, if we reduced our performance indicator set from the outset, we would not account for the inherent selection bias in our survey data. As we have seen in this analysis, selection bias does not exist uniformly across the indicators, but varied depending on which outcome indicator was considered. Secondly, if we reduced the dimensions from the outset by, say, only using the indicators for employment, overall satisfaction, and post-training salaries, we would forego a sizeable amount of information that would be lost in the process of data reduction. In the specific analysis we have presented here this loss of variance would amount to 20% (table 2.42). Additionally, from the viewpoint of policymakers, institutional researchers, and other stakeholders, it may be more desirable to have a larger array of indicators to enable more fine grained investigations into institutional differences. Thirdly, computationally it is not significantly more arduous to create eight indicators over three. However, it

remains important to be mindful of the underlying dimensionality when evaluating the performance indicators developed in this study.

Another item of interest was if the institutions studied could be grouped into a manageable number of strata showing similar performance characteristics and whether these strata could be related back to other external demographic background characteristics that were not used to create the performance indicators. To do this, we employed a hierarchical cluster analysis using Ward's minimum variance method with squared Euclidean distance as the dissimilarity measure. In the cluster analysis of institutions, we used the eight performance indicators developed in this study and standardized them. The resulting dendrogram can be seen in figure 2.9.

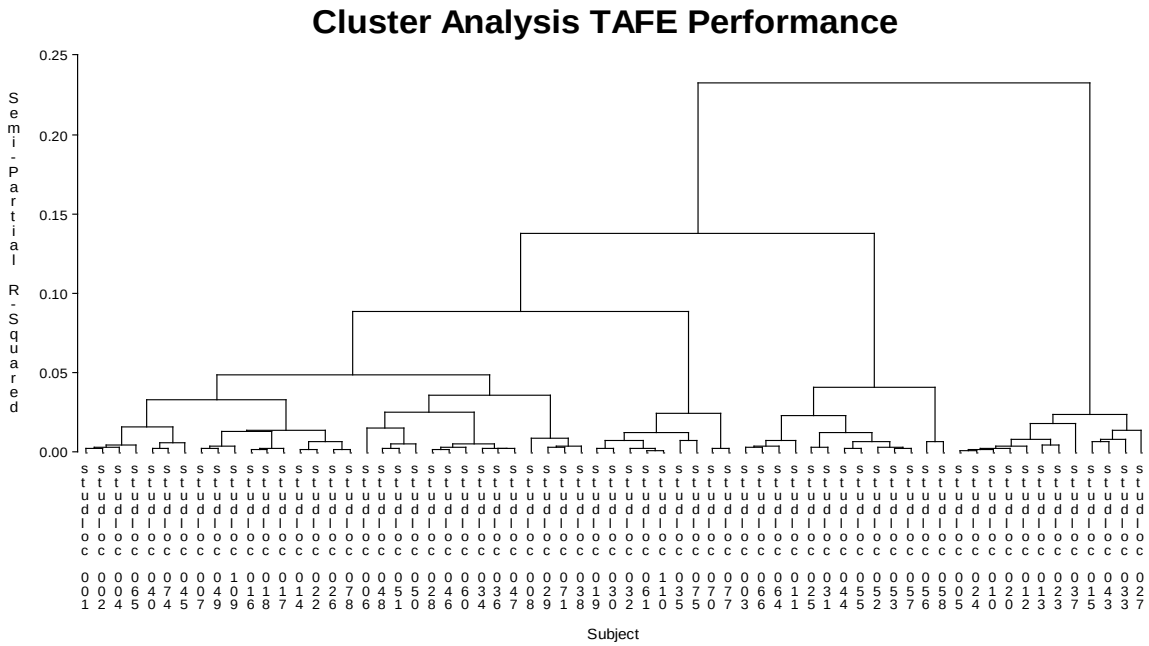

Figure 2.9 Cluster dendrogram resulting from all eight indicators

Selecting the appropriate type and number of clusters from a hierarchical cluster analysis can be difficult (see, for instance, Hartigan, 1985). We have been guided by the quest for a manageable number of clusters of around four to six and employed the cubic clustering criterion as suggested by Sarle (1983)

as the determining factor for the number of clusters. The graph of the cubic clustering criterion by number of clusters is displayed in figure 2.10.

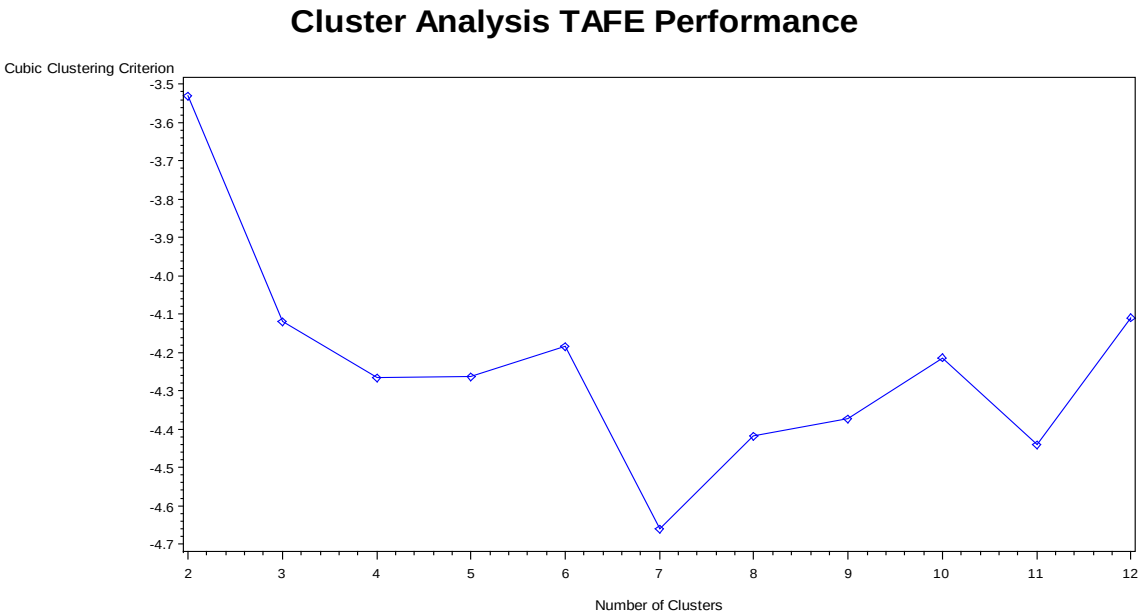

Figure 2.10 Cubic clustering criterion by number of clusters

It can be seen that for Ward's method, the cubic clustering criterion has a peak at six clusters within the range of clusters we were aiming to consider.

To assess the performance characteristics of the resulting clusters we created the mean scores of all standardised individual indicators and graphed them against our three performance components (figure 2.11).

The effect of post training salaries was the most striking feature of this graph. While practically all other clusters indicated salaries of around zero (for example, the national average without cluster 6) salaries in cluster 6 were about three standard deviations above the mean. In terms of the employment/achievement component, cluster 1 rated highest, with clusters 5 and 6 also well above the mean. Clusters 5 and 6 also performed superior on student satisfaction, whereas cluster 2 scored about one or more standard deviations below the mean. Overall, the six clusters ranked differently across

the eight indicators with some key reversals appearing on the assessment satisfaction, salary and general satisfaction items.

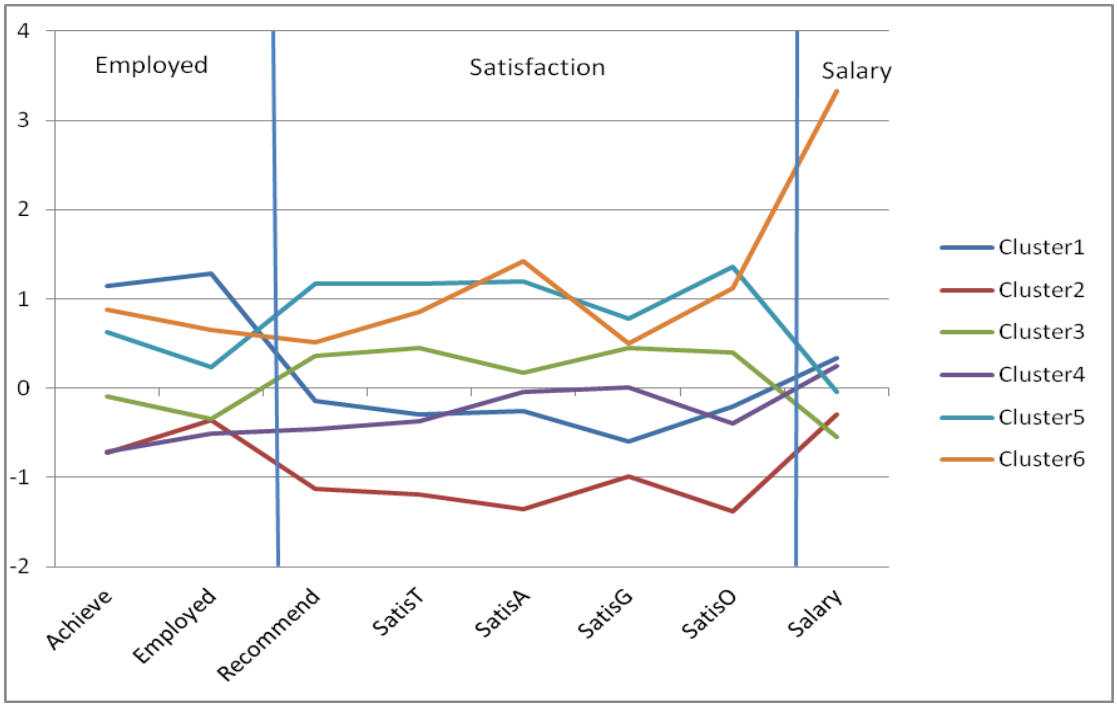

Figure 2.11 Performance profiles of all indicators against clusters

We were also interested in assigning a composite performance score to each cluster. Creating a summary performance measure and thus allocating an 'ultimate' ranking of the performance of TAFE institutes is somewhat problematic. Firstly, the performance measures created in this study were based on the student outcome survey, and thus reflected the types of measurements that could be extracted from that survey. Secondly, we believe that the validity of a summary measure of performance ultimately depends on the preferences of the user of such a measure. For instance, a policymaker may be inclined to put higher weight on the 'objective' measures (for example, employment and salary outcomes) analysed in this study, whereas a researcher with a psychological or sociological bent could be more interested in a summary measure which puts higher weight on satisfaction outcomes. We therefore considered several methods of creating a summary performance indicator. A very basic method would be to average all eight institutional

scores. Another method worthy of consideration would be to extract the main component of all eight indicators via principal component analysis and then use the factor scores of the first component as a measure of a summary score. In our example, a summary score derived by this method would account for almost half of the variance of all individual indicators (see table 2.42), and could thus be considered a fairly adequate representation of all eight measures. However, as we have seen in table 2.44, it is mostly the satisfaction items which load on this first component, and therefore using first factor scores would bias the overall performance score strongly toward our satisfaction scores.

To alleviate this bias problem we standardised all indicators and determined the maximum value for each indicator. The resulting eight maximum values could then be conceptualised as the coordinates of an 'ideal performance point' (IPP) in an 8-dimensional space.

$$\text{IPP} = (\text{Employed}_{max}, \text{Salary}_{max}, \ldots\ldots, \text{Achieve}_{max}) \qquad (13)$$

If there was an institute that would score the maximum on all eight indicators, that institute would represent the IPP. We then calculated the distance of each institute from the IPP in this multi-dimensional space via the generalised Pythagorean theorem for Euclidean distance:

$$Inst_i = \sqrt[2]{(Empl_{max} - Empl_{inst(i)})^2 + (Sal_{max} - Sal_{inst(i)})^2, \ldots, (Ach_{max} - Ach_{inst(i)})^2} \qquad (14)$$

The resulting score for each institute represents a 'distance from maximal performance' score for that institute, where a lower score signals a better performance profile. These scores were then used to rank all institutes. Table 2.45 displays the institutional 'distance from maximal performance' scores in descending order by institution. It can be seen that institute 31 emerged as the top performer, followed by institutes 56 and 55. At the low

performance end we find institute 37 and, after a considerable gap, institutes 10 and 8.

Table 2.45 Summary 'distance from maximal performance' score by institution

Institute	Score	Institute	Score	Institute	Score	Institute	Score
31	3.15	32	5.09	34	5.85	23	7.34
56	3.40	16	5.13	60	5.87	50	7.43
55	3.44	7	5.19	19	5.91	33	7.43
44	3.59	61	5.23	1	5.97	6	7.47
25	3.64	77	5.32	22	5.98	13	7.69
58	3.69	26	5.39	75	6.01	5	7.80
57	3.83	28	5.42	14	6.04	20	7.89
3	3.96	78	5.43	30	6.06	38	8.06
11	4.22	47	5.52	4	6.11	43	8.12
66	4.30	65	5.56	109	6.25	12	8.22
53	4.42	40	5.60	45	6.26	24	8.25
64	4.49	110	5.72	36	6.26	27	8.39
52	4.65	49	5.73	48	6.95	15	8.40
70	4.68	35	5.76	71	6.97	10	8.95
17	4.79	74	5.77	29	7.18	8	9.08
2	4.91	46	5.80	51	7.21	37	10.96
18	5.04						

The institutional scores displayed in table 2.45 were then used to calculate cluster performance scores. Table 2.46 shows cluster membership of institutions with similar performance.

Table 2.46 Cluster membership of similarly performing institutions

Cluster	Mean distance from maximal performance score	Institutions in cluster
1	5.53	061, 110, 032, 070, 077, 019, 030, 035, 075
2	8.29	005, 024, 010, 020, 012, 013, 023, 015, 043, 033, 027, 037
3	5.59	014, 022, 026, 078, 016, 018, 001, 002, 007, 049, 040, 074, 017, 004, 109, 065, 045
4	6.79	028, 046, 036, 047, 048, 051, 034, 029, 071, 060, 038, 050, 008, 006
5	3.97	053, 057, 044, 055, 003, 066, 052, 025, 031, 064, 011
6	3.55	056, 058

It is apparent that clusters 5 and 6 encompass the high performing institutions (having the closest distance to the maximal performance), whereas cluster 2 comprised institutes from the bottom end of the institutional ranking. The cluster mean performance scores enabled us to relate these scores to the

variables 'Indigenous', 'remoteness', and 'SES percentile', in an effort to uncover patterns that may be of interest to policymakers. We limited the analysis to these three key characteristics, as most of the other demographic variables of interest were used to construct the performance scores. As these three variables all displayed different characteristics, each of the variables required a different type of analysis. SES percentile had the properties of an interval variable. We performed a one-way ANOVA to determine whether there were significant differences between the clusters based on SES (table 2.47).

Table 2.47 SES percentile means by cluster

Cluster	Students	Mean SES percentile	Std dev
1	8,197	50.5	77.1
2	15,510	63.4	93.9
3	12,334	45.2	84.4
4	17,166	52.4	81.3
5	9,574	41.4	71.7
6	966	47.9	84.4
	F=1178.74 Pr<0.001		

While the F test in the above analysis indicated that there were differences between the clusters in terms of SES percentile, it was not clear which clusters differed significantly from each other. We therefore added a Bonferroni posthoc test to determine significance. The outcome is displayed in table 2.48.

Table 2.48 Significance of differences of SES percentile by cluster

	Cluster					
Cluster	1	2	3	4	5	6
1						
2	***					
3	***	***				
4	***	***	***			
5	***	***	***	***		
6	-	***	-	***	***	

Note: *** indicates significance at 0.05

It can thus be seen that every cluster was different from each other cluster, with the exception of clusters 3/6 and 1/6. The results were somewhat

unexpected and thus interesting. Clusters 5 and 6, the two lowest ranking clusters by SES are the two highest ranking clusters by institutional performance. Inversely, the lowest performing institutes represented by clusters 2 and 4 were also the clusters with the highest SES. While it would be difficult to derive direct policy implications from this finding, it does challenge the often seen conjecture that lower SES levels correlate with poorer social outcomes. There is no way of ascertaining whether or not this relationship is causal; however, these results show that being located in (or attended by students from) lower socioeconomic background areas was no impediment to strong institutional performance for TAFE institutes, at least in the context of the performance measures analysed in this study.

We were also interested in evaluating whether there was a relationship between the percentage of Indigenous students and performance cluster membership. As seen in the previous analysis, institutes with student body backgrounds from lower socioeconomic backgrounds performed rather well on our performance scales. It was therefore worthwhile to test if a similar pattern applied to clusters with higher percentages of Indigenous students as there is often a strong inverse correlation between the proportion of Indigenous individuals and SES (see Hunter 1996). We performed a Chi square test of the weighted student population to see whether there were any differences with respect to cluster membership and proportion of Indigenous students within individual clusters. The resulting chi square test and cross-tabulations are displayed in table 2.49. The specific test performed was a Rao-Scott test, which was an adjusted Pearson Chi square test (see Rao & Scott 1981) that took into account the complex stratified design features of the Student Outcomes Survey.

Table 2.49 Frequencies, percentages, and chi square test of Indigenous by cluster

Cluster		Indigenous	Non-Indigenous	Total
1	Frequency	3,668	60,305	**63,973**
	Row	5.7	94.3	
2	Frequency	3,125	164,391	**167,516**
	Row	1.9	98.1	
3	Frequency	9,307	132,228	**141,534**
	Row	6.6	93.4	
4	Frequency	4,616	134,462	**139,078**
	Row	3.3	96.7	
5	Frequency	5,671	83,576	**89,247**
	Row	6.4	93.7	
6	Frequency	909	5,478	**6,387**
	Row	14.2	85.8	
Total		**27,295**	**580,439**	**607,734**
		4.5	**95.5**	**100**
	Rao-Scott ChiSquare		445.846	
	DF		5	
	Pr > ChiSq		<0.0001	

The results of the analysis of the relationship between institutes and the proportion of Indigenous students were somewhat similar to the analysis of SES. The highly significant Chi square indicated that there was a significant relationship between Indigenous student numbers and clusters. Specifically, cluster 2, representing the bottom end of the performance scale, had the lowest proportion of Indigenous students, whereas the highest performing cluster, number 6, comprised by far the largest Indigenous percentage. Similar to the SES analysis, this was a finding that was somewhat unexpected and therefore noteworthy.

Finally, we performed a non-parametric test to see whether there were differences based on the remoteness of students' residence and cluster membership. To do this we recoded individual students' ARIA category (ABS 2013b) of geographic location into an ordinal variable, beginning with '1' for major city, '2' inner regional, '3' outer regional, '4' remote, and '5' very remote. This coding scheme yielded a variable with descending urbanity from 1 to 5 which can then be averaged over institutions and clusters of institutions. Subsequently we calculated Wilcoxon rank sum scores and performed a test

devised by Kruskal and Wallis (1952) to ascertain whether clusters differed by distribution of the remoteness variable. Table 2.50 displays the results.

Table 2.50 Wilcoxon scores and Kruskal-Wallis Chi square test: clusters by remoteness

Cluster	N	Mean score	Sum of scores	Expected under H0	Std dev under H0	Mean score
1	8,152	2.4	345019759	258247208	1409081.47	42323.3266
2	15,443	1.2	320218258	489218797	1806807.41	20735.4955
3	12,217	1.9	418358853	387022343	1660265.18	34243.9922
4	17,082	1.5	461401464	541140678	1867486.13	27010.9744
5	9,497	2.4	407666585	300855463	1502246.52	42925.8276
6	966	4.3	54421485.5	30601914	515661.07	56336.9415
			K-W Chi-Square		17938.6	
			DF		5	
			Pr > Square		<0.0001	

It was clear from the Kruskal-Wallis Chi square in table 2.50 that the clusters differed significantly by degree of urbanity/remoteness. A more rural inclination was noticeably related to clusters exhibiting higher performance, as evidenced by the means of 2.4 and 4.3 for clusters 5 and 6. Conversely, the clusters containing the lowest performing institutes, cluster numbers 2 and 4, were also the clusters displaying the highest degree of urbanity.

It is not the purpose of this study to determine causal relationships between demographic, environmental, and other endogenous variables and the effectiveness of TAFE institutes. Furthermore, it can be argued that the items investigated here — indigeneity, remoteness, and socioeconomic background — were significantly inter-related. It may also be possible that the growth in mining in recent years explains part of the strong performance of the more remote institutions. This would certainly be a reasonable hypothesis as far as some of the objective outcomes (such as employment and salaries) investigated in this research are concerned. In any event, this brief analysis of three key student body background variables revealed that there were significant associations between such characteristics and TAFE performance which may warrant further analysis.

2.1.21 Modelled versus raw indicators

After considering the performance indicators introduced in this research one might ask: Why go through such complicated algorithms and calculate these complex modelled indicators if one could simply resort to plain indicators, such as straight means or percentages of, say, post-training employment? After all, some government-sponsored evaluation tools do just that (see, for instance, *myuniversity.gov.au* for the use of unadjusted raw student outcome data).

In earlier sections we have already alluded to the limitations of this approach. These are mainly that such unadjusted indicators do not account for the specific demographic and educational profiles of individual institutions. Neither do they adjust for employment and salary conditions in the residential area of students. Also, these indicators do not take into account any selection bias that may have occurred in the process of data collection.

We believe we have thus made a strong case for the application of the complex methodology proposed in this research to create such indicators. There is however one question that remains: To what extent do our modelled and adjusted results differ from simple raw scores? We will consider this question by focusing on the employment indicator developed in section 1.15.

First we calculated the mean employment rate for all institutes considered in this analysis and then determined the difference between an individual institute's employment rate and the overall mean. The resulting values can be found in the 'Raw' column in table 2.50. The modelled difference from the mean can similarly be found by first determining the overall mean

$$\mu_{overall} = \frac{e^{intercept}}{1+e^{intercept}} \qquad (15)$$

and then the institutional mean

$$\mu_{institute} = 1 - \frac{e^{-(intercept+instituteestimate)}}{1+e^{-(intercept+instituteestimate)}} \qquad (16)$$

with the modelled score then derived by

$$Modelled = \mu_{institute} - \mu_{overall} \quad (17)$$

Table 2.51 displays differences between modelled and raw income scores by institution. The second (and sixth) column contains the actual difference between the raw overall and institutional post-training employment rate. The third (and seventh) column contains the results from our model, for example, the difference to the overall mean when covariates are taken into account. The fourth (and eighth) column displays the difference between the 'raw data' and 'modelled' columns in percentage points.

While the correlation between modelled and raw measures is about 0.75 ($p < 0.001$), there were some significant differences between the two methods. We can see that out of our 65 institutions, in 16 (24.6%) the difference between modelled/raw scores was between 5 and 10 percentage points, and in one further institute (1.5%) the difference exceeded 10 percentage points. It was thus evident that over a quarter of all institutes differed by five or more percentage points with respect to the average score.

Table 2.51 Differences between modelled and raw income scores by institution

Institute	Raw data	Modelled	Difference	Institute	Raw data	Modelled	Difference
1	-2%	-3%	1%	36	4%	1%	3%
2	-4%	-1%	-3%	37	-7%	-6%	-1%
3	-4%	0%	-4%	38	-8%	-6%	-2%
4	-10%	-7%	-3%	40	4%	-1%	6%
5	-5%	-3%	-2%	43	-4%	-1%	-3%
6	-3%	-3%	0%	44	5%	4%	1%
7	0%	1%	-1%	45	-7%	-7%	0%
8	-16%	-10%	-6%	46	-2%	4%	-6%
10	-6%	-4%	-2%	47	8%	0%	8%
11	-2%	-3%	1%	48	0%	0%	0%
12	-7%	-2%	-5%	49	8%	2%	6%
13	2%	4%	-1%	50	0%	-5%	5%
14	2%	0%	2%	51	-2%	-1%	-1%
15	4%	4%	0%	52	4%	-1%	5%
16	7%	2%	5%	53	4%	0%	4%
17	4%	0%	4%	55	4%	3%	1%
18	5%	0%	5%	56	8%	0%	9%
19	7%	6%	1%	57	6%	2%	4%
20	-1%	-2%	1%	58	14%	5%	9%
22	0%	0%	-1%	60	10%	2%	8%
23	-2%	-3%	0%	61	-3%	4%	-8%
24	-2%	-4%	2%	64	4%	2%	1%
25	6%	0%	6%	65	-1%	0%	-1%
26	1%	-2%	4%	66	3%	-1%	4%
27	2%	1%	1%	70	13%	8%	4%
28	5%	3%	1%	71	-15%	-6%	-9%
29	-8%	-5%	-3%	74	-4%	-1%	-3%
30	5%	7%	-1%	75	4%	3%	1%
31	7%	4%	4%	77	7%	5%	2%
32	0%	2%	-3%	78	-2%	-2%	0%
33	-3%	1%	-4%	109	-14%	0%	-14%
34	0%	0%	0%	110	5%	3%	1%
35	8%	3%	4%				

Note: In the shaded areas, difference is > = +/- 5 percentage points.

An alternative way of visualising the differences between raw and modelled performance indicators is to display the data in a two-dimensional space (figure 2.12). If adjusting for all the different covariates as described in previous chapters had no effect, then we would see a straight line with a 45 degree slope in the graph. Instead, the extent of variation between modelled and raw values can clearly be seen. This underscores the necessity to employ a well-fitting model over a simple calculation of raw scores to construct such performance indicators.

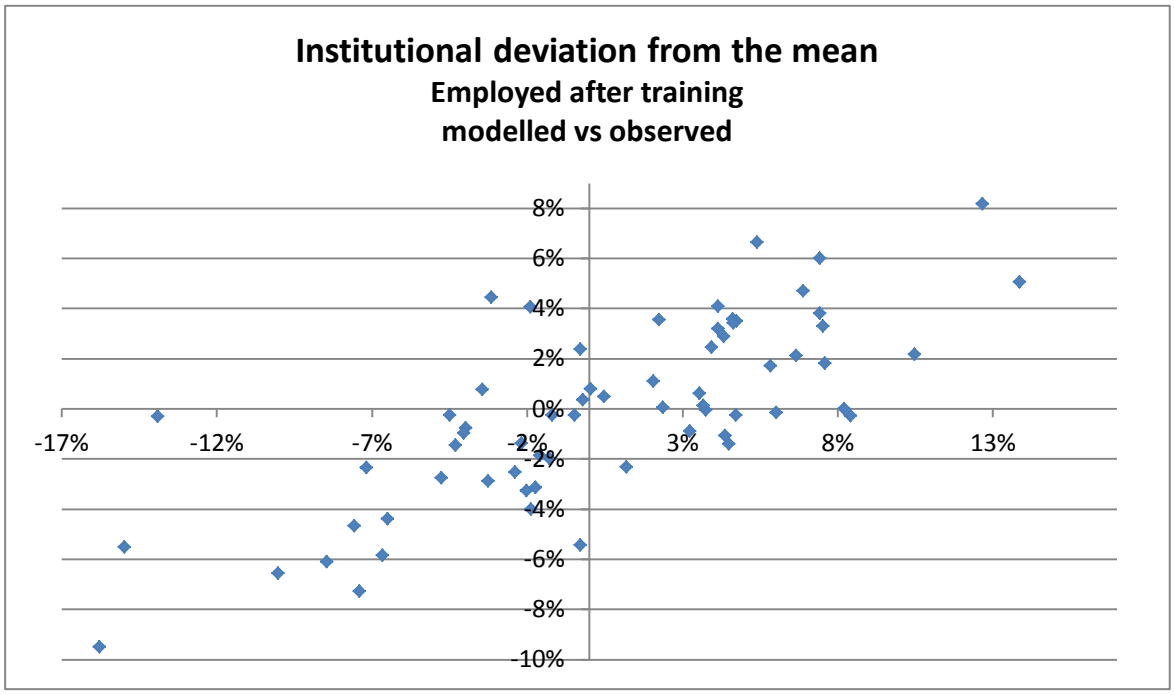

Figure 2.12 Differences between raw and modelled performance indicators

2.1.22 Conclusion

In this analysis we have seen that it was possible to use Student Outcomes Survey data to create performance indicators which were then used to quantify institutional differences based on those indicators. In order to account for different student demographics, institutional profiles and post-training labour market conditions, we adjusted our analyses for a variety of covariates, such as student age, student sex, field of education, qualification level, graduate/module completer status, labour force status prior to training and employment and wage conditions in the area of student's residence.

The research presented in this study makes a number of contributions to academic research in vocational education. First, a methodology for creating and applying performance indicators using primarily the Student Outcomes Survey was devised, statistically validated and applied to the data for 2011. This methodology allowed us to go far beyond the usage of conventional performance measures (such as 'percentage employed' or 'proportion of satisfied students' etc.). The techniques proposed here enabled us not only

to account for differences between institutions in respect to demographic and educational student characteristics, but also to quantify the extent to which institutions differed based on the variance that could be attributed to inter-institute variation. Additionally, we have adjusted some of the models with data external to the SOS, such as employment data and income data from the residential area of respondents, in order to further improve the validity of our indicators.

Second, the response rate to the Student Outcomes Survey 2011 was relatively low, which gave reason to suspect that the key outcomes investigated here could be subject to selection bias. In this research, we detected that several of our performance indicators did exhibit selection bias, namely post-training salary, teaching and overall satisfaction, perception of achievement and willingness to recommend the institution of training. With the selection models applied here we have been able to quantify selection bias and deployed measures to adjust our outcomes for it. The primary source of selection bias found among the performance indicators investigated was respondent age.

Third, the institutional comparison tables gave a clear picture of how TAFE institutes compared in regard to individual performance indicators. Those that scored significantly below or above the national average could be easily visualised via a shading scheme or alternatively via caterpillar plots. In addition, we grouped similarly performing institutes into performance clusters and related institutional effectiveness to some environmental characteristics, such as socioeconomic background, urbanity and proportion of students with Indigenous background. A surprising finding was that higher institutional effectiveness was associated with lower SES, a larger degree of remoteness and larger proportion of Indigenous students. This may indicate that these populations particularly benefit from the vocational education sector.

Fourth, a further important finding was that the modelled performance indicators proposed in this paper appeared to have a significant advantage over simple (raw) indicators. Modelled performance indicators enabled us to compare institutions on a more even footing as we were adjusting for the most important demographic and environmental variables. As a result we found that, while the correlation between raw and modelled performance indicators was quite strong, there were significant differences between the scores of individual institutions.

Finally, we found that the proposed eight performance indicators could be condensed into three coherent components: student satisfaction and willingness to recommend, achievement and employment outcomes, and post-training salary. While it would be possible to create these three indicators from the raw survey data via methods such as principal component analysis, one should keep in mind that some of the original eight performance indicators contained selection bias and the three summary components could therefore only be created after the problem of selection bias had been addressed.

We should note that, as in any study, there are some limitations to this analysis that should be kept in mind when interpreting the results. Firstly, the models employed here obviously capture only those covariates we were able to obtain from our secondary sources. In particular, the models relating to student satisfaction, while statistically significant, had only limited explanatory power. This meant there were other effects that impacted on student satisfaction which were not captured within the Student Outcomes Survey. Moreover, all our models indicated that the majority of differences between institutes was due to individual differences rather than to the institutions themselves. What we have essentially done in this research is shine a light on the efforts to determine differences between the institutions.

A second limitation was the relatively poor response rate to the SOS, which was in the area of 35% for the 2011 cohort. This issue was further exacerbated in the model analysing post-training salaries as only a third of SOS respondents answered the salary question. As in any parametric statistical analysis, with a smaller number of observations came reduced statistical power.

Finally, as in most surveys, there was some inherent respondent bias in the SOS. This was often caused by higher (or lower) response rates of particular groups (in this specific research, most of the response bias effects could be attributed to the age of the respondent). While we have attempted in this analysis to employ corrective measures that aimed to alleviate (non-) response bias, the possibility cannot be excluded that some residual survey bias exists.

There are some potential implications for policymakers that emerge from this research. First, it is possible to develop both a profile for institutional performance as well as a defensible single index for ranking — neither should be used without the other; second, institutional performance is multidimensional, which may demand a multidimensional policy response; and third, the results of this study call into question possible beliefs by policymakers and the media that lower performing institutions are associated with the 'poorer' sections of Australian society and that perhaps policy intervention to lift institutional performance might be best focused elsewhere (for instance, on more urban institutions).

The research presented here introduced a methodology to produce eight performance indicators using the Student Outcomes Survey and some census and labour market data. At this present time the lack of availability of additional data for more potential indicators limits the applicability of this methodology. However, in the course of the current drive by the Federal Government for more transparency in the VET sector (Transparency agenda, 2012), it can be expected that in the near future more institutional data will

be collected and it may be possible to create additional performance indicators. It will then be beneficial if validated methodologies, such as developed in this study, will be available.

3 On the efficiency of Australian TAFE institutes: Portfolio paper 2

3.1 Structure

3.1.1 Introduction

Increased competition for public funding has made governments, tax payers and other stakeholders more aware of the need for educational institutions, like TAFE institutes, to demonstrate improved productivity. Efficiency of TAFE institutes is thus of great interest to policymakers, regulators, consumers and to the institutions themselves. Knowledge about institutional efficiency may be useful to government agencies in allocating funds and in assessing the impact of funding decisions on the relationship between financial and administrative input into institutions versus the produced output, for instance in form of hours taught, graduates produced, employment effects and other outcomes. Institutions could use information about their own efficiency to benchmark themselves against other institutions and to make adjustments to their own resource allocation. Regulators could also use this knowledge to potentially identify areas of high risk in the delivery of vocational education and training (VET). Moreover, due to limited market mechanisms in the

provision of educational products, knowledge of alternative means for establishing benchmarks of efficiency is of importance to all stakeholders in educational institutions. Due to the usefulness of efficiency indicators in funds allocation, process improvements and input and output optimization, research involving efficiency analysis has elicited a growing body of academic literature in recent decades. The contemporary approach to the analysis of the efficiency in the form of a production function was pioneered by Farrell. In his seminal paper, Farrell (1957) argued that the measurement of efficiency was necessary to ascertain whether additional inputs were needed to increase outputs or if outputs could be increased by raising efficiency alone. He also developed a generalisable production function which enabled the computation of efficiency measurements under multiple input scenarios. In the 1970s two teams independently developed two different techniques enabling the measurement of efficiency. Aigner, Lovell and Schmitt (1977) formulated the first stochastic frontier model (SFA) in the form of a parametric maximum likelihood technique which overcame previous limitations of frontier estimation by partitioning the error term into a normal 'noise' component and a one-sided inefficiency component. Almost at the same time, Charnes, Cooper and Rhodes (1978) presented their work on a non-parametric linear programming method called data envelopment analysis (DEA). Their method focused on a scalar measurement of the efficiency of each unit under consideration, obtained after determining appropriate weights from the observed data for multiple inputs and outputs. While data envelopment analysis is well-suited to the calculation of efficiencies in the context of multiple inputs and outputs, stochastic frontier analysis in its basic form is best suited to one output in the presence of multiple inputs.

The introduction of both of these methods stimulated a host of applied research. One of the features of production frontier analyses is their utility in multiple input and output scenarios, which makes this type of efficiency

analysis particularly useful for non-commercial entities where traditional accounting methods provide little benefit in establishing the efficiency of such units (in the DEA framework often referred to as decision making units [DMU]). While both methods have been used in the analysis of commercial contexts (see, for instance, Cullinane et al., 2002 for an SFA of port efficiency or Fraser & Cordina, 1999 for a DEA of agricultural efficiency), one of their main applications has been the efficiency analysis of public institutions and government-owned entities. The spectrum of sectors analysed has varied across a wide field of institutional units, ranging from hospitals (see, for instance, Biorn, 2003), public transport (for example, Karlaftis ,2004), public utilities (for example, Pahwa et al., 2003) and prisons (for example, Butler & Johnson 1997), to numerous applications to educational contexts.

In this study we will employ parametric SFA and non-parametric DEA to estimate the efficiency of Australian TAFE institutes. We will review the theoretical underpinnings of both techniques and the results from both quantitative methods will be compared and analysed in respect to their differences. We will look into the statistically significant drivers of efficiency and assess whether (and if so, which) specific policy recommendations could be derived from such analyses. An additional aspect of this study will be whether there are economies of scale effects in the TAFE sector. We will thus analyse to what extent institutional size is a determinant of efficiency. While our main aim is to examine the performance of TAFE institutes and to explain why performance varies between units, we also aim to find answers to the following questions:

- What inputs and outputs provide reliable means to determine the efficiency in the Australian TAFE sector?
- What are the strengths and weaknesses of the two techniques (stochastic frontier analysis and data envelopment analysis) with respect to efficiency

measurement of vocational education institutions? To what extent are the results arising from both techniques related?

- How is institutional efficiency distributed within the ATS and which environmental variables help to predict patterns of efficiencies and why?
- What could be done to improve performance?

In this study, we will proceed in the following manner: Firstly, we will review the theoretical underpinnings of the techniques used. Secondly, we will identify the appropriate variables that will be used in the respective analyses. Next, we will describe the available data and the way they were collected. Then we will evaluate predictors of institutional efficiency, compare results of both techniques and discuss differences. Finally, we will consider what practical relevance the research results have and identify any concrete policy implications emerging from our findings.

3.1.2 SFA and DEA and their applications in education

Since the late 1970s, Stochastic Frontier Analysis and Data Envelopment Analysis have been increasingly used in the estimation of production frontiers. While the two methods aim to quantify the relative efficiency of a set of economic units under consideration, both use quite different approaches to arrive at their respective results. In fact, these results may differ somewhat substantially even if identical input and output measures are used (see, for instance, Becalli et al. 2006 or Fiorentino et al. 2006). It is thus necessary to consider some basic differences between the two methods. The SFA approach has, as the name suggests, a stochastic underpinning, meaning that the computation of the maximum likelihood estimation rests on an underlying assumed error distribution and is therefore parametric. DEA, on the other hand, produces a deterministic frontier. As such, DEA is much more sensitive to individual DMUs and specifically those which are instrumental in establishing the location of the production frontier. DEA is inherently better suited to deal

with multiple input and output scenarios, while special accommodations need to be made when applying SFA to multiple output scenarios. As a parametric method, SFA relies on an explicit specification of a mathematical production function, whereas DEA as a deterministic procedure does not require such a specification. In DEA, the efficiency frontier is determined by those DMUs that exhibit the highest efficiency (those DMUs consequently attain the top efficiency of 1.0), whereas in SFA the frontier is estimated based on all available data points and no single DMU must necessarily have an efficiency value of 1.0.

Efficiency analysis utilising DEA or SFA has been applied frequently in educational contexts. Worthington (2001) reviewed the use of efficiency frontier analysis in the education sector. He found increasing usage of efficiency analysis particularly in Europe and North America. There have been some applications at the secondary school level (for instance, see Kirjavainen & Loikkanent, 1998 for a DEA or Mizala et al., 2002 for a SFA example). One British study (Lenton, 2006) analysed colleges of further education in the UK with respect to their efficiency compared with US colleges. This study was notable in that it was a rare incidence of efficiency frontier analysis being applied in the technical and further education sector. Lu and Chen (2013) appraised the efficiency of technical universities and vocational education institutions in Taiwan. They used frontier analysis and found comparative disadvantages in efficiency of vocational institutes by contrast with technical universities. However, in general, most of the empirical applications of these methods have been confined to universities (see, for instance, Kempkes & Pohl, 2010; Taylor & Harris, 2004; Hanke & Leopoldseder, 1998).

Despite the popularity of econometric frontier analysis overseas, the existing published research utilising SFA or DEA in Australian educational contexts is somewhat limited. Academic inquiry into the efficiency of secondary schools is

scant. Bradley et al. (2004) utilised data from a subset of Queensland schools to perform a DEA with respect to a number of school- and teacher-specific input variables and numeracy- and literacy-related output variables. To date, this remains the only published study applying efficiency frontier analysis in Australian secondary education. There is a more sizeable body of published research involving universities. Avkiran (2001) used 1995 data from Australian universities in a DEA to determine universities' efficiency with respect to the delivery of educational services and fee-paying enrolments. Other analyses examining cross-sectional university performance were performed by Abbott and Doucouliagos (2003), Carrington et al. (2005), Horne and Hu (2008), Worthington and Lee (2008) and Abbott and Doucouliagos (2009). Abbott and Doucouliagos (2009), for example, carried out a comparative SFA of New Zealand and Australian universities with respect to the impact of the proportion of foreign students on institutional efficiency. Their analysis indicated that intensified competition for overseas students resulted in increased efficiencies in Australian universities, while such competition had no impact on New Zealand universities. Finally, only a small number of studies involving Australian TAFE institutes could be identified. These were notably the research by Abbott and Doucouliagos (1998 and 2002), who performed DEAs utilising data only from Victorian institutes, and one nationwide DEA study by Fieger et al. (2010). These three studies used a two-step efficiency analysis process where, in the first stage, the efficiency of all units were estimated using DEA, and then, in a second stage, Tobit regression was applied to determine the drivers of efficiency. The results of the Victorian and nationwide analyses showed a broad dispersion of efficiencies. Abbott and Doucouliagos found a wide range of predictors of institutional efficiency, including class size, contact hours, proportion of executive staff, direct delivery expenditure, staff training, capacity utilisation, proportion of students with disability and government funding. Fieger's study of nationwide

TAFE institutes employed a slightly different set of independent variables in the Tobit regression and identified institute remoteness as the main predictor of efficiency. There has been no previous efficiency analysis of the Australian TAFE sector which utilised the stochastic frontier approach. Consequently, there has been no investigation which simultaneously evaluates the two methods of DEA and SFA in the Australian TAFE context. A comparative application of both approaches was thought to assist in elucidating their respective benefits and disadvantages in the particular environment of vocational education system as well as to provide a means for cross-validating the results of applying one method with the outcomes of the other method. The present study aimed to address these issues and thus to fill a gap in applied economic educational research.

3.1.3 Stochastic Frontier Analysis

The foundations for this methodology were laid by Aigner, Lovell and Schmidt (1977) who formulated the first stochastic frontier model. Their main contribution was the introduction of a new approach to the specification of the error term, namely its partitioning into a normal 'noise' component and a one-sided inefficiency component. Stochastic frontier production functions are an extension of the classic Cobb-Douglas (1928) function which can generally be expressed in this form:

$$Y = e^{\beta_0} X_1^{\beta_1} X_2^{\beta_2} \dots X_n^{\beta_n} e^{\varepsilon} \quad (1)$$

This model can then be transformed and linearised by taking the log of both sides:

$$ln(Y) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) \dots \beta_n \ln(X_n) + \varepsilon = \beta_0 + \sum_{i=1}^{n} \beta_i \ln(X_i) + \varepsilon \quad (2)$$

This notation is easily recognisable as a variation of the classical multiple regression model in which *Y* stands for the output, $X_{1\ to\ n}$ for the vectors of input, β_o for the intercept, $\beta_{1\ to\ n}$ for the beta weight estimates of the vector's relative contribution to the prediction model, and ε for statistical noise. This

model, being a regression model, means that it computes a predicted output for a given configuration of inputs and individual observed outputs *Y* will deviate randomly above and below the calculated regression surface, thereby contributing to ε. The contribution of Aigner et al. (1977) was to postulate that the error term ε actually comprised two error components, one being the statistical noise portion *v*, and the other being the non-negative technical inefficiency *u* which is distributed independently from *v.*

$$\varepsilon_i = v_i - u_i \tag{3}$$

The original Cobb-Douglas function can thus be re-formulated as

$$ln(y_i) = \beta_0 + \sum_{i=1}^{n} \beta_i \ln(X_i) + v_i - u_i \tag{4}$$

where technical inefficiency TE_i can then be determined by

$$TE_i = e^{-u_i} \tag{5}$$

TE_i is meant to be located between 0 and 1 and is ordinarily assumed to be positively half-normally distributed[1]. Aigner et al. determined the mean of ε and *u* as:

$$\mu_\varepsilon = \mu_u = -\sigma_u \sqrt[2]{\frac{2}{\pi}} \tag{6}$$

and the variance of the error ε as:

$$var(\varepsilon) = var(u) + var(v) = \frac{\pi-2}{\pi}\,\sigma_u^2 + \sigma_v^2. \tag{7}$$

where σ_u represents the variance of the normal distribution prior to truncation to 0. The total variance in the error term is given by σ_ε^2.

$$\sigma_\varepsilon^2 = \sigma_v^2 + \sigma_u^2 \tag{8}$$

The ratio of the standard deviation of the inefficiency component to the standard deviation of the 'noise' error component is given by lambda:

$$\lambda = \frac{\sigma_u}{\sigma_v} \tag{9}$$

and gamma γ is an indicator of the portion of the one-sided error component in the overall variance.

$$\gamma = \frac{\sigma_u^2}{\sigma_\varepsilon^2} \tag{10}$$

[1] In this study we will apply the half normal error assumption, which is the most frequently applied distribution (Cullinane 2006). There are, however, other error distribution assumptions possible, such as exponential, half-truncated, gamma etc.

These simple relationships provide convenient means to assess the quality of the results of a SFA. For instance, if $\lambda \rightarrow 0$ this implies that $\sigma_v^2 \rightarrow \infty$ and/or $\sigma_u^2 \rightarrow 0$ which indicates that symmetric error (that is, 'statistical noise') dominates the overall error component. Similarly, if $\lambda \rightarrow \infty$ then $\sigma_u \rightarrow \infty$ or $\sigma_v \rightarrow 0$ whereby deviation from the frontier can be explained by inefficiency. Following from this, when $\gamma \rightarrow 1$ the inefficiency component increasingly accounts for error in the model.

3.1.4 Data Envelopment Analysis

Non-parametric Data Envelopment Analysis was first proposed by Charnes, Cooper and Rhodes (1978). Their proposed model was specifically designed to enable the evaluation of the efficiency of not-for-profit entities, which they termed 'decision making units' (DMU). The model utilises linear programming to provide a scalar measure of the efficiency of each participating DMU. One particular feature that makes this technique attractive is the possibility of analysing units that have multiple inputs and outputs.

DEA obtains the efficiency measure of individual outputs in the form of a non-linear programming formulation as the maximum of the ratio of the weighted outputs to the weighted inputs (11) under the condition that the same ratio of all other units is ≤ 1.

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r \; y_{r0}}{\sum_{i=1}^{m} v_i \, x_{i0}} \qquad (11)$$

on the condition that

$$\frac{\sum_{r=1}^{s} u_r \; y_{rj}}{\sum_{i=1}^{m} v_i \, x_{ij}} \leq 1$$

Here, u_r and v_i (u_r, $v_i \geq 0$) are the variable weights to be calculated by the solution of (11) and y_{rj} and x_{ij} (y_{rj}, $x_{ij} > 0$) are the outputs and inputs of the j_{th} DMU. Similarly, inefficiency can be defined by

$$\min f_0 = \frac{\sum_{r=1}^{m} v_i \; x_{i0}}{\sum_{r=1}^{s} u_r \, y_{r0}} \qquad (12)$$

on the condition that

$$\frac{\sum_{i=1}^{m} v_i \; x_{ij}}{\sum_{r=1}^{s} u_r \, y_{rj}} \geq 1$$

where y_{rj} and $x_{ij}$are the known outputs and inputs of the j_{th} decision-making unit and u_r and $y_r \geq 0$ are the variable weights to be determined by the data on all of the decision-making units in the reference set. This setup can be conceptualised by considering a simple one-input, one-output scenario with three decision-making units (figure 3.1, adapted from Banker et al. 1984, p.1079). The production function specifies the maximum possible output (*y*) that can be achieved with a given input (*x*).

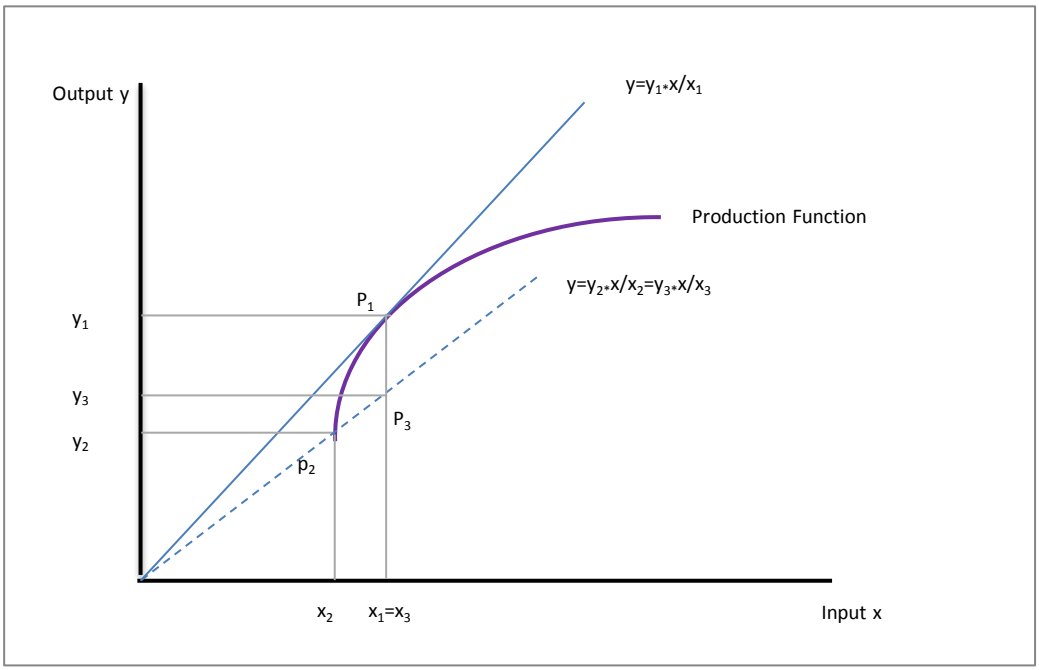

Figure 3.1 Schematic one input/one output DEA scenario

Here, DMUs P_1 and P_2's location on the production frontier indicates that they achieve maximum output with their given inputs. P_3 on the other hand is inefficient as it fails to produce the same output as P_1, despite having the same input, and therefore is not located on the production function.

It is further notable that the DEA framework allows for the analysis of DMUs under different 'returns to scale' scenarios. The initial method, as proposed by

Charnes, Cooper & Rhodes (1978) uses the constant returns to scale approach. This method, often referred to as 'CCR', assumes that changes in inputs to DMUs lead to proportional changes in output. This method was later refined by Banker, Charnes and Cooper ('BCC', 1984) to also allow for a variable returns to scale setting. Under variable returns to scale assumptions, DMUs may also produce disproportionally lower outputs as inputs increase (for example, decreasing returns to scale) or, conversely, disproportionally greater outputs as inputs increase (for example, increasing returns to scale).

While we have omitted a multiple input and output scenario in this small example, the major advantage of DEA over SFA is that it deals easily[2] with such multiple input and output settings. DEA differs from SFA in that it is a deterministic method, which as a non-parametric technique does not require a specification of a functional form or the consideration of underlying distributions. While this property of DEA can be advantageous, it also means that there are no straightforward[3] statistical hypothesis tests. DEA is also sensitive to proper selection of the data. Theoretically it is possible that two individual outlier DMUs define the production frontier, which could lead to distorted calculated efficiencies for all other units. DEA is also prone to determining a significant number of DMUs under consideration as efficient which may diminish the utility of the method. This is because as the number of DMUs increases, the chance of encountering a DMU close to the true production frontier increases, and therefore the frontier determined by DEA approaches the true frontier asymptotically as the number of units increases (Banker 1989). Notwithstanding this finding, Zhang and Bartels (1998) have shown that the average technical efficiency of firms decreases as the number of firms increases.

---

[2] There are ways to adapt SFA to a multiple input/output setting via a distance function; however, this requires transformations of input and outputs and has implications for the interpretation of some results. For an example, see Newman and Matthews (2006).

[3] Some approaches have been developed to address this issue. For a discussion of these approaches see Banker 1996.

3.1.5 Data characteristics and preparation

The aim of this study is to ascertain the efficiency of Australian TAFE institutes via SFA and DEA and to determine what exogenous variables drive the calculated efficiencies. When deciding on an approach to undertake efficiency frontier analysis in the TAFE sector one has to bear in mind that it is necessary to take into account some specific circumstances that are unique to that sector. Similar efficiency frontier analyses involving universities or secondary schools can often rely on data such as number of full time staff, staff qualifications, number of graduates, test scores, grades, research outputs such as publications and conference presentations, successful grant applications and others. Data comparable to the aforementioned are difficult to obtain for TAFE institutes. There is obviously a scarcity of research and research-related inputs and outputs that relate to TAFE institutes. Many TAFE institutes employ a large percentage of part-time lecturers, this proportion differs from TAFE to TAFE and no reliable data about this proportion can be obtained. Many TAFE institutes do not award grades for some or all of their courses but rely on 'competency-based' assessments.

It can therefore be seen that there are some contextual factors that constrain the specification of frontier efficiency models for the TAFE sector. The majority of those factors can be categorized into three groups:

- the total absence of data for the entire sector (for example, staff qualifications)
- partial data only available for a subset of TAFE institutes (for example, certain types of financial data)
- data that are too dissimilar in nature from TAFE to TAFE due to the lack of a comprehensive national reporting standard (for example, assessment beyond competency-based assessment).

Despite the aforementioned difficulties we have been able to assemble and derive a dataset containing the necessary information to undertake the course

of research set out in earlier paragraphs. The data used in this study came from a variety of sources. These sources included institutional annual reports, information on institutional websites, the Student Outcome Survey and the Students and Courses database at NCVER. Of significance was the choice of year(s) for which data should be obtained. It was intended to assemble a panel of data comprising a number of years in an effort to:

- maximize the number of data points
- enable analysis of changes in efficiency over a given period.

However, data collection was more difficult than anticipated as institutes do not publish financial data in a uniform pattern. Specifically, the collecting of several consecutive years of financial data appeared to be difficult. It was thus decided to focus on one particular year with the following stipulation:

- the year had to be as recent as possible
- it had to be an augmented SOS year[4] to enable the use of the most robust institutional data
- the chosen year had to have the maximum number of available data points.

Taking these considerations into account, 2011 was chosen as the year of analysis.

It was initially planned to use the following financial variables in our analysis: salary and related expenditure, depreciation and amortization expenses, supplies and services expenditure, other expenses and total expenses. A number of institutes, notably all TAFE institutes in Western Australia and Victoria, plus several others post these figures contained in their annual reports on their websites. We approached all remaining institutes or state regulatory authorities with the request to make their respective data available

[4] Odd years feature an augmented sample of the SOS, containing about 300,000 questionnaires with an approximate 33% response rate. In these years the SOS is designed to enable estimates at an institutional level. In even years the SOS sample contains about 100,000 questionnaires, and the focus of estimates is the state level.

for this study. These requests were only partially successful and several institutes/authorities decided not to participate in this research. Other institutes/authorities submitted only partial data, mostly limited to total expenditure. A number of institutes released their data to the author on the condition that they would not be individually identified in this research.

The initial plan was to include all 69 Australian TAFE and TAFE-like institutions[5] in this analysis. However, this intention was impeded by a number of factors. In addition to those institutes who did not provide data, some institutions proved to be too specialised to be compared on an equal footing with the majority of TAFE institutes. These were notably the Driver Education Centre of Australia and the National Art School. Some of the TAFE units in dual-sector universities did not have segregated financial data for their TAFE branch available. After considering availability of data for the remaining institutes it was decided to include those units in the final data set which had data for the total expenditure variable in 2011 available. This yielded 56 TAFE institutes for inclusion in the analysis.

In addition to financial expenditure data, various other variables used in the efficiency analysis were derived from NCVER's Students and Courses database. These included the hours of delivery. This variable indicated the number of student contact hours by institution. We also used the hours of delivery variable to derive an additional variable, funding per hour of delivery. This variable represented the ratio of total expenditure to hours of delivery.

In an effort to create a proxy indicator of institutional completion rates we also queried the Students and Courses database. As there were no officially published or endorsed completion rates and some known difficulties with their

[5] In the context of this study, the term 'TAFE and TAFE-like institute' refers to TAFE institutes, TAFE divisions of a university, skills institutes and polytechnics, from here on only referred to as 'TAFE'.

establishment existed (see, for instance, Mark & Karmel 2010 and Ross 2011)[6], we devised a simple scheme to approximate a completion rate. For this, we extracted the number of course completions (*C*) and the number of course enrolments in certificates I, II, III, IV and diplomas from the database and reconstructed the overall enrolment (*ER*) for all five qualifications (*i*) by institution. We did this for two consecutive years (*t-1*) and (*t*) to alleviate the effect of qualifications that had not been completed within a single year. The proxy for the institutional completion rate (*CR*) was then given by (1):

$$CR = \frac{C_t + C_{t-1}}{\sum_{i=1}^{5} ER_t + \sum_{i=1}^{5} ER_{t-1}} * 100 \qquad (13)$$

In the absence of uniformly reported grades achieved by students during their studies, we calculated a proxy, which took into account the competency-based assessment nature of the TAFE system. This proxy was called the load pass rate (LPR) and was an indicator of student success rates during their studies. The LPR should not be confused with the completion rate, as the former dealt with modules and the latter with courses. The LPR was defined as the ratio of hours (or full-year training equivalents [FYTEs]) attributed to students who gained competencies/passed assessment in an assessable module or unit of competency to all students who were assessed and either passed, failed or withdrew. The calculation was based on the annual hours (or FYTEs) for each assessable module or unit of competency and included competencies achieved/units passed through recognition of prior learning. The load pass rate was calculated using the number of competency achieved *(A)*, recognition of prior learning granted *(RPL)*, competency not achieved *(F)* and withdrawals *(W)* via the following formula (ANR 2011, p.208):

$$LPR = \frac{A+RPL}{A+RPL+F+W} * 100 \qquad (14)$$

[6] These difficulties mostly stem from the absence of a unique student identifier in the Australian vocational education system and variances in the duration of a number of VET qualifications. A detailed discussion of this issue can be found in paper 3 of this portfolio.

A number of other variables were sourced predominantly from the 2011 SOS. These included institutional proportions in terms of sex, student type (module completers/graduates), Indigenous students, students who used a language other than English at home, disabled students and the proportion of certificate III or higher of each institute's graduates. Other variables included were the average age of the student body at individual institutions, an average institutional remoteness score derived from the ARIA[7] variable, an average institutional student satisfaction score based on the relevant 5-item satisfaction scale contained in the SOS, as well as a proxy for the average socioeconomic background of students attending an institution, derived from individual SEIFA[8] scores of students. We also used the SOS to calculate the number of different courses offered by each institution which had at least one student enrolled[9]. An ordinal categorical variable indicating size was derived from the total expenditure variable. The categories created were 'very large', signifying expenditure in excess of $120,000,000, large ($70,000,000 to $120,000,000), medium ($45,000,000 to $69,999,999), small ($25,000,000 to $44,999,999), and very small with expenditure smaller than $25,000,000.

Finally, we utilised one variable which indicated the institutional employment rate of students after their training. This variable was determined via a mixed effects regression model in which we adjusted for a number of covariates in an effort to make institutes more comparable with respect to demographic and environmental variables. All data needed for this derived variable were sourced from the Student Outcome Survey 2011. Specifically, we entered the variables of the unemployment rate in the residential area of students and the institution itself as random effects, and field of education, employment status prior to the training, student age, sex, qualification level enrolled in,

[7] Using the ABS ARIA definition, (see ABS 2013b).

[8] Socio Economic Indexes For Areas (SEIFA) is a product developed by the Australian Bureau of Statistics that ranks areas in Australia according to relative socioeconomic advantage and disadvantage. The indexes are based on information from the five yearly Census. For more information see ABS (2013).

[9] This requirement was meant to reflect the administrative effort required by an institute to run a course.

unemployment rate and graduate or module completer status as fixed dummy-coded effects into this model. A more thorough description of the methodology applied can be found in section 2.1.15 of paper 1 of this portfolio.

3.1.6 SFA 'Teaching hours' model

In this study we evaluated the efficiency of a number of TAFE institutes. To do this we specified two conceptually different models. On one hand, our interest was in determining institutional efficiency based purely on basic financial expenditure and administrative input and the produced output as measured by teaching contact hours. This model was meant to measure predominantly the simple conversion of financial resources into teaching output irrespective of how 'successful' the teaching was. We called this model the 'teaching hours' model. Then we developed a second model. This model evaluated efficiency based on a tangible student outcome measure, namely the adjusted employment outcome by institutes. The idea here was that we considered the attainment of employment after students completed their training as one of the main societal objectives of the TAFE system and that it was therefore prudent to evaluate if resources were used efficiently to attain this goal. In this model we employed institute-specific variables such as funding per teaching hour, completion rates, load pass rates, proportion of students enrolled in certificate III or higher, proportion of graduates over module completers and the size of the respective institute as predictors. This second model was called the 'employment model'.

The reason for this dual-model approach was simple: The 'teaching hours' model was very mechanistic and assessed efficiency without any regard to quality of the achieved output. It thus purely evaluated efficiency as a 'cost of service' provision. The 'employment model' on the other hand incorporated a degree of quality assessment into the efficiency analysis, as its output measurement went beyond the simple provision of service and included how

beneficial the training was in securing employment. It could therefore be argued that the employment model evaluated efficiency (for example, in terms of utilisation of inputs to achieve certain output) as well as effectiveness (for example, the quality of the output achieved).

The starting point for operationalising the teaching hours model was a production function as expressed by the Cobb-Douglas equation:

$$T = e^{\beta_0} E^{\beta_1} C^{\beta_2} e^{\varepsilon} \qquad (15)$$

where *T* denoted the output in teaching hours, *E* the total expenditure and *C* the number of courses offered by a given TAFE. C was included as it was an indicator of the complexity of college administration. Taking the natural logarithm of both sides of (15) and accounting for the SFA specific error component (Battese & Coelli, 1995) created the following transformation:

$$\ln(T_i) = \beta_0 + \beta_1 \ln(E_i) + \beta_2 \ln(C_i) + v_i - u_i \qquad (16)$$

Descriptive statistics for the variables used in estimating the production frontier can be found in table 3.1

Table 3.1 Descriptive statistics 'Teaching hours' model variables

Variable	N	Mean	StdDev	Minimum	Maximum
Teaching hours	56	5,521,177.5	4,174,682.5	473,279	22,346,943
Total expenditure	56	79,966,968.0	53,563,163.2	12,324,312	288,974,000
Number of courses offered	56	172.6	83.3	32	439

In addition to the frontier production function in (16) we intended to investigate which exogenous variables may be influencing the teaching hours technical efficiency. We therefore specified a second component in which we included some variables which were hypothesised to influence efficiency:

$$\mu = \delta_0 + \sum_{k=1}^{K} \delta_k z_k \qquad (17)$$

Here, *z* represented the hypothesised *K* predictors of efficiency and δ the parameters that needed to be estimated. In our 'teaching hours' model we hypothesized that predominantly demographic factors influenced efficiency, as

these factors likely required administrative adjustments to TAFE operations. We therefore entered the variables with institutional indicators for English as a second language, disability, remoteness, age and sex into our efficiency model (for descriptive statistics see table 3.2).

Table 3.2 Descriptive statistics for 'teaching hours' efficiency predictor variables

Variable	N	Mean	Std dev	Minimum	Maximum
English second language	56	16.3	9.8	4.6	40.2
Students with disability	56	9.4	2.9	4.4	18.5
Remoteness (ARIA)	56	2.1	1.0	1.1	4.7
Student age	56	33.0	2.2	27.6	37.1
Proportion of males	56	57.2	10.7	32.8	96.6

This type of scenario would have originally been estimated in a two-step approach, where the first step specified the stochastic production frontier and led to the estimation of efficiency scores and the second step estimated the relationship between efficiency scores and efficiency predictors. Wang and Schmidt (2002) demonstrated that this two-step procedure was biased and that instead stochastic frontier models and the way in which efficiency u_1 depended on predictors could be estimated in a single step using maximum likelihood estimation.

Analysis by Waldman (1982) showed that for the specification of a stochastic frontier model it was beneficial to examine the third moments of the least squares residual. If this quantity was positive, then the least squares slope estimates and $\lambda = 0$ represented a local maximum of the likelihood. Conversely, if the third moment was negative, the likelihood had a greater value at some other point where $\lambda = 0$. Waldman demonstrated that a positive third moment reduced the slope of the maximum likelihood estimation to the slope of OLS regression, and consequently there are no inefficiencies and all DMUs are located on the efficiency frontier. On the other hand, negative skewness of the residuals of the OLS regression warranted that maximum likelihood estimation was indeed the appropriate procedure to estimate the production frontier.

We thus embarked on our analysis with the formulation of a linear regression model identical to our proposed SFA model. The results can be seen in table 3.3 (model 1). The third moment based of the OLS residuals was estimated to be -0.63, thus signalling appropriateness of maximum likelihood estimation of the stochastic frontier. While the coefficients of the OLS model only had limited usefulness, they provided a meaningful starting point for the maximum likelihood estimation (Cullinane et al., 2006). The *R*-squared estimate of the OLS was a fairly substantial 0.91 and indicated that most of the variation in teaching hours could be explained by total expenditure and number of courses offered by institute. The two independent variables were highly significant and both exhibited the sign that would be expected, for example, higher expenditure and increasing number of courses tended to be associated with a rise in teaching hours.

We then estimated our basic stochastic frontier model, using the same variables (table 3.3, model 2). While coefficients and intercept had the same sign as in OLS regression, along with similar magnitude and strong significance, the real interest here was in the estimated variance parameters.

Table 3.3 'Teaching hours' model estimates

Variables	OLS		MLE			
	Model 1		Model 2		Model 3	
	Est	*P>|t|*	*Est*	*P>|z|*	*Est*	*P>|z|*
Stochastic frontier						
Constant	-4.221	<.001	-4.022	<.001	-2.730	<.001
Total expenditure	0.926	<.001	0.989	<.001	0.968	<.001
Number of courses offered	0.553	<.001	0.345	<.001	0.134	0.025
Inefficiency model						
Constant					-17.631	0.001
English second language					0.129	0.027
Students with disability					0.053	0.726
Remoteness (ARIA)					2.708	<.001
Student age					-0.074	0.768
Proportion of males					0.112	0.048
R-squared	0.913					
Wald Chi-sq			385.4	<.001	983.5	<.001
σ_v			0.126	<.001	0.127	<.001
σ_u			0.387	<.001		
σ^2			0.165	<.001		
Lambda			3.073	<.001		
Gamma			0.904			

The strong significance of the Wald test indicated that the coefficient(s) were significantly different from zero and thus confirmed the model's explanatory power. The variances of the random error and inefficiency components of the model, σ_u and σ_v, were both significant. The significance of λ confirmed the presence of inherent statistical inefficiency in the data. The estimate for gamma was quite high at 0.9 and denoted that 90% of the variability in delivered teaching hours could be attributed to technical inefficiencies. The closeness of γ to 1 pointed towards the existence of a deterministic production frontier (Parsons, 2004). The significance of γ and λ affirmed the preponderance of inefficiency in the composite error term and also validated SFA as the appropriate tool for this specific analysis (Chen, 2007). We will investigate in the next section whether this meant that individual efficiency results calculated by SFA correlated strongly with results obtained from DEA. Additionally a test was performed to determine whether the DMUs investigated by our Cobb Douglas model used constant-returns-to-scale technology. This hypothesis tested whether the coefficients in the model added up to 1. The

sum of the coefficients for 'total expenditure' and 'number of courses' was calculated as 1.33 and the test for equality to 1 yielded a chi-squared value of 6.54 ($p = 0.0106$), therefore we rejected the hypothesis of constant returns to scale technology with the value greater than 1 indicating an increasing returns to scale setting. In the scenario considered, this meant that outputs would increase disproportionally when inputs were increased.

Having gained insights into the characteristics of our basic frontier model we proceeded to specify the SFA model that included explanatory variables for the technical inefficiency variance function (table 3.3, model 3). First we noted that parameters and significance of the frontier function were comparable to the model without the inefficiency terms. The Wald chi-squared and variance component of the random error term of the whole model were also significant and of similar magnitude. The main items of interest in model three were thus the inefficiency effects. We noted that the proportion of students with a disability and the institutional mean age of the student body were not related to institutional efficiency. The strong significance of remoteness as indicated by individual students' ARIA classification pointed to inefficiency being a function of remoteness. This result confirmed the findings of Fieger et al. (2010), who found remoteness to be the key variable predictive of inefficiency. The finding of remoteness-related inefficiency may be partially attributed to the vast distances between major centres and the associated difficulties with transport and infrastructure in remote areas. However, it must be pointed out that 'remoteness' acts also as a proxy for institution size as urban institutes tend to be larger than rural institutes. Internationally, remoteness was rarely identified as a driver of inefficiency in post-secondary education, although Izadi et al. (2002) found some incidental relationship between remoteness and inefficiency. In model 3 we found further, albeit weaker, positive predictive associations between the proportion of males and inefficiency and the proportion of students with English as a second language

and inefficiency. Possible explanations here may be that males tend to be engaged at higher rates in apprenticeships, which require larger administrative and financial efforts on the part of the institution. Larger financial, educational and administrative efforts may also be at play when considering the relationship between increasing inefficiency and higher rates of non-native English speakers. An assessment of the correlation between the proportion of males and the proportion of apprentices and trainees in the 2011 data revealed a value of 0.44 ($p<0.001$), thus supporting this explanation. Similarly, larger proportions of students with English as a second language may necessitate more intensive teaching modes, such as lower teacher/student ratios, which may in turn explain some variation in institutional inefficiency in respect to the percentage of non-native English speakers.

3.1.7 SFA employment model

After evaluating the teaching model and its fairly basic inputs and output we were interested in the specification of the production function of our second model, for example, the employment model. The principal conceptual difference from the first model was the capturing of a different type of efficiency that included what may be considered a 'quality component' as our output. As such, one could argue that this model dealt not only with efficiency per se, but also with institutional effectiveness. Such an approach was not unique in educational efficiency analysis. Mizala et al. (2002) drew on grades in a DEA of Chilean schools, Kanep (2005, cited in Perreira & Moreira, 2007) utilized drop-out rates in a study of Estonian secondary schools in a DEA study, Oliviera and Santos (2005) applied the free disposal hull methodology[10] and used school attendance rates and Perreira and Moreira (2007) employed test scores as their respective output variable. In the Australian TAFE system there was no similar variable that was available across all institutions to enable SFA.

[10] Free disposal hull methodology is another tool for production frontier analysis, developed by Deprins, Simar, and Tulkins (1984).

There was no uniform grading system across all Australian TAFE institutes and neither were test scores or attendance rates available. We therefore decided on an output which incorporated the assessment of one of the core purposes of the national vocational education system: the gaining of employment after the completion of the training. This employment output variable was described earlier in section 2.1.15; however, it is worth re-emphasizing that this variable was adjusted for a number of demographic and environmental covariates in order to enable inter-institutional assessment of this outcome on an even footing. The adjustment of this variable therefore included covariates such as the local unemployment rate in the residential area of the respondents as well as a number of other educational and demographic variables.

To specify the employment model we proceeded through the same sequence of stages as the teaching hours model. First we specified the production frontier model, then we conducted an OLS regression to get an initial idea about the fit of the model and to determine the third moments of the residuals, then we ran the basic SFA model and then the full SFA model including the hypothesised heteroskedastic inefficiency effects. The basic Cobb-Douglas production function for the employment output was formalized as:

$$E = e^{\beta_0} F^{\beta_1} CR^{\beta_2} C3^{\beta_3} G^{\beta_4} e^{S^{\beta_{5-8}}} e^{\varepsilon} \qquad (18)$$

where *E* signified the institutional employment score, *F* the funding per teaching hour, *CR* the completion rate, *C3* the proportion of certificate III or higher graduates, *G* the proportion of graduates over module completers, and *S* a five-level dummy coded categorical variable, indicating the size of the institution. Transforming (18) into a function for the estimation of the stochastic production function produced:

$$\ln(E_i) = \beta_{i0} + \sum_{k=1}^{6} \beta_{ik} \ln(x_{ik}) + v_i - u_i \qquad (19)$$

Descriptive variables to determine the stochastic production frontier are shown in table 3.4[11]

Table 3.4 Descriptive statistics 'Employment outcome' model

Variable	N	Mean	Std dev	Minimum	Maximum
Employment outcome	56	0	0.14	-0.31	0.33
Funding per hour	56	17.93	11.83	8.92	87.30
Completion rate	56	27.31	11.81	4.60	72.20
Certificate III or higher	56	82.08	8.68	52.32	96.71
Graduates	56	38.94	14.33	15.73	76.22
Very small	56	0.07	0.26	0	1
Small	56	0.25	0.44	0	1
Medium	56	0.21	0.41	0	1
Large	56	0.23	0.43	0	1
Very large	56	0.23	0.43	0	1

For the heteroskedastic inefficiency component we hypothesised a number of variables which could be expected to impinge on efficiency. These were predominantly the demographic variables from our first model, for example, age, sex, disability, remoteness and English language status, as well as all the predictors from the production function (for descriptive statistics, see tables 3.2 and 3.4).

The starting point for this model was again an OLS regression model (table 3.5, model 1). The *R*-squared value for the OLS employment model was 0.30, a value considerably smaller than in the 'teaching hours' model. Coefficients of the predictor variables displayed some unexpected properties. Only the proportion of graduates was significant at the 95% level. A higher proportion of graduates was associated with a lower employment score. Another interesting result was that funding per teaching hour was not related to employment outcomes. With respect to institutional size, compared with very large institutions, medium and smaller institutions had marginally significant superior employment outcomes. We calculated the third moment of the

[11] Prior to using the 'Employment outcome' variable in the production function it was re-centred so that the original lowest value of 0.306 (the original minimum) was added to all employment outcome values.

residuals of the OLS model as -0.54. This negative skewness validated the intended SFA approach.

Table 3.5 Employment outcome model estimates

Variables	OLS		MLE			
	Model 1		Model 2		Model 3	
	Est	*P>|t|*	*Est*	*P>|z|*	*Est*	*P>|z|*
Stochastic frontier						
Constant	-0.167	0.800	0.285	0.651	0.228	0.836
Funding per hour	0.010	0.828	0.001	0.976	0.018	0.711
Completion rate	-0.025	0.477	-0.030	0.309	-0.008	0.923
Cert III or higher	0.190	0.154	0.107	0.384	0.030	0.872
Graduates	-0.092	0.024	-0.077	0.014	-0.008	0.794
Very large	-	-	-	-	-	-
Large	0.046	0.227	0.051	0.100	0.052	0.172
Medium	0.078	0.053	0.089	0.005	0.152	0.003
Small	0.073	0.084	0.077	0.014	0.052	0.103
Very small	0.050	0.401	0.023	0.638	0.007	0.917
Inefficiency model						
Constant					-1.610	0.871
English second language					0.078	0.112
Students with disability					0.416	0.061
Remoteness (ARIA)					2.233	0.004
Student age					-0.493	0.076
Proportion of males					-0.003	0.944
Funding per hour					-0.044	0.331
Completion rate					0.048	0.495
Cert III or higher					0.041	0.642
Graduates					0.017	0.684
Load pass rate					-0.025	0.735
Very large					-	-
Large					1.333	0.272
Medium					1.320	0.286
Small					-0.689	0.685
Very small					-4.861	0.275
R-squared	0.302					
Wald Chi-squared			28.080	<0.001	22.870	<.001
Sigma v			0.037	0.001	0.043	<.001
Sigma u			0.131	<0.001		
Sigma 2			0.018	<0.001		
Lambda			3.510	<0.001		
Gamma			0.925	<0.001		

Model 2 (table 3.5) represented the basic SFA model without inefficiency effects. Variances of the idiosyncratic (σ_v) and inefficiency (σ_u) components were significantly different from 0. The γ value of 0.92 pointed to the

existence of a deterministic frontier and the significance of λ denoted the presence of inefficiency. The test for the hypothesis of constant returns to scale technology was performed by determining the sum of the coefficients. This summation yielded 0.24 (chi-squared for difference from one was 16.43 [p<.001]) which suggested that TAFE institutes under this model operated under a decreasing returns to scale environment. This meant that when inputs were increased under this scenario, outputs increased at a lower rate than inputs.

The full SFA model including inefficiency effects can be found as model 3 in table 3.5. Parameter estimates and slope signs of this model were comparable to the basic SFA model. The inefficiency component of the model indicated that remoteness was strongly associated with inefficiency. This replicated the main result of the 'teaching hours' model, which also ascertained remoteness as a key predictor of inefficiency. Two additional inefficiency predictors exhibited marginal significance[12]. These included the proportion of students with a disability and average age of the student body. Students with disabilities may have greater difficulty in obtaining post-study employment which could contribute to lower employment outcomes and thus explain why higher proportions of them appear to be associated with lower employment efficiency. The average age of the student body was negatively related to inefficiency. We hypothesised that this result may be due to the higher propensity of more mature students to achieve superior employment outcomes. To test this we used SOS 2011 data and ran a logistic regression model with employment outcome as the dependent variable and a dummy-coded age group variable as the sole predictor. The result (table 3.6) showed that most of the age groups had a much higher likelihood of being in post-training employment compared with 15 - 19 year olds. However, this may also be due to more mature students already being in pre-training employment.

[12] In this portfolio paper, we consider a p-value of 0.05 <p <0.10 'marginal'.

Table 3.6 Odds ratios for employment status after training by age group

Age group vs 15 to 19 years	N	Odds ratio	Conf interval	
20 to 24 years	13,433	1.903	1.802	2.01
25 to 29 years	7,593	2.145	2.006	2.294
30 to 34 years	6,832	1.99	1.859	2.132
35 to 39 years	7,449	2.203	2.058	2.358
40 to 44 years	8,272	2.414	2.258	2.58
45 to 49 years	8,309	2.784	2.599	2.983
50 to 54 years	7,469	2.65	2.469	2.844
55 to 59 years	5,409	2.335	2.161	2.523
60 to 64 years	3,026	1.359	1.245	1.482
65 years and over	1,388	0.512	0.458	0.572
Wald Chi-sq = 2233.5389 P >ChiSq <0.001				

However, other studies (Daghbashyan, 2012; Nisantha & Ranasinhe, 2012) querying the impact of student age on educational efficiency did not find a similar relationship.

After verifying the suitability of our two models and discussing the interpretation of model statistics and coefficients it is now of interest to consider the actual estimated efficiencies of individual institutions. The efficiencies followed from (5) and for the half-normal production model were specifically derived by

$$TE = \{\frac{1-\Phi(\sigma_* - \mu_{*i})}{1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)}\}\exp(-\mu_{*i} + \frac{1}{2}\sigma_*^2) \qquad (20)$$

where Φ signified the cumulative distribution of the normal distribution and μ_{*i} and σ_* were defined as

$$\mu_{*i} = -\epsilon_i \sigma_u^2 / \sigma_s^2 \qquad (21)$$

and

$$\epsilon_i = y_i - x_i \beta \qquad (22)$$

and

$$\sigma_* = \sigma_u \sigma_v / \sigma_s \qquad (23)$$

The calculated efficiencies for the ‘teaching hours’ and ‘employment outcome’ model are shown in table 3.7.

Table 3.7 Estimated SFA efficiencies for 'Teaching hours' and 'Employment outcome' models

Institute	Technical efficiency	
	Teaching hours	Employment outcome
1	0.984	0.909
4	0.977	0.820
5	0.973	0.927
7	0.932	0.950
10	0.953	0.870
11	0.943	0.860
13	0.971	0.989
14	0.953	0.859
15	0.978	0.982
16	0.966	0.991
17	0.953	0.981
18	0.986	0.944
19	0.960	0.973
20	0.963	0.973
22	0.862	0.976
23	0.921	0.924
24	0.964	0.878
25	0.968	0.980
26	0.908	0.896
27	0.985	0.990
28	0.973	0.939
29	0.959	0.871
30	0.987	0.978
31	0.967	0.992
32	0.866	0.968
33	0.982	0.956
34	0.996	0.973
35	0.920	0.929
36	0.986	0.891
37	0.979	0.719
38	0.991	0.780
40	0.621	0.946
43	0.960	0.926
44	0.946	0.941
45	0.893	0.669
46	0.980	0.983
47	0.916	0.955
48	0.927	0.963
49	0.992	0.983
50	0.972	0.819
51	0.981	0.954
52	0.739	0.938
53	0.840	0.995
55	0.967	0.978
56	0.474	0.932
57	0.723	0.997

Continued next page

Institute	Technical efficiency	
	Teaching hours	Employment outcome
58	0.327	0.953
60	0.389	0.995
64	0.948	0.969
65	0.977	0.940
66	0.979	0.947
70	0.918	0.986
71	0.978	0.724
74	0.198	0.885
77	0.983	0.983
110	0.423	0.994
Mean	0.888	0.929
SD	0.182	0.074

While the average efficiencies of both models were fairly similar (0.89 and 0.93), the efficiencies of the 'teaching hours' model appeared to encompass a far greater range than the 'employment outcome' efficiencies. Histograms of both efficiencies can be seen in figures 3.2 and 3.3.

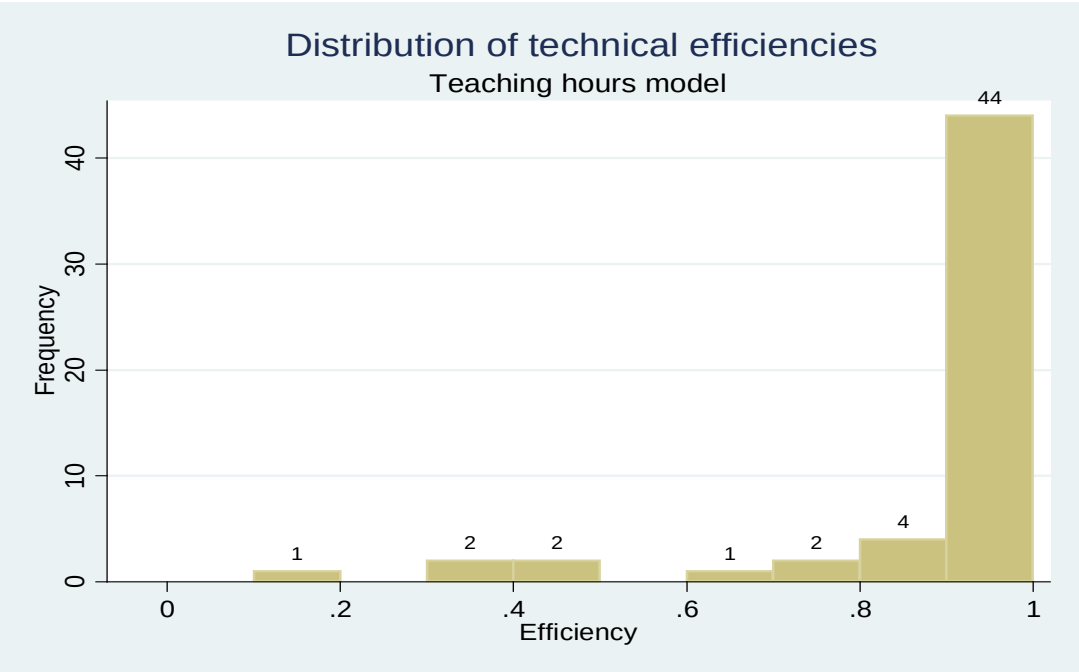

Figure 3.2 Distribution of SFA efficiencies: Teaching hours model

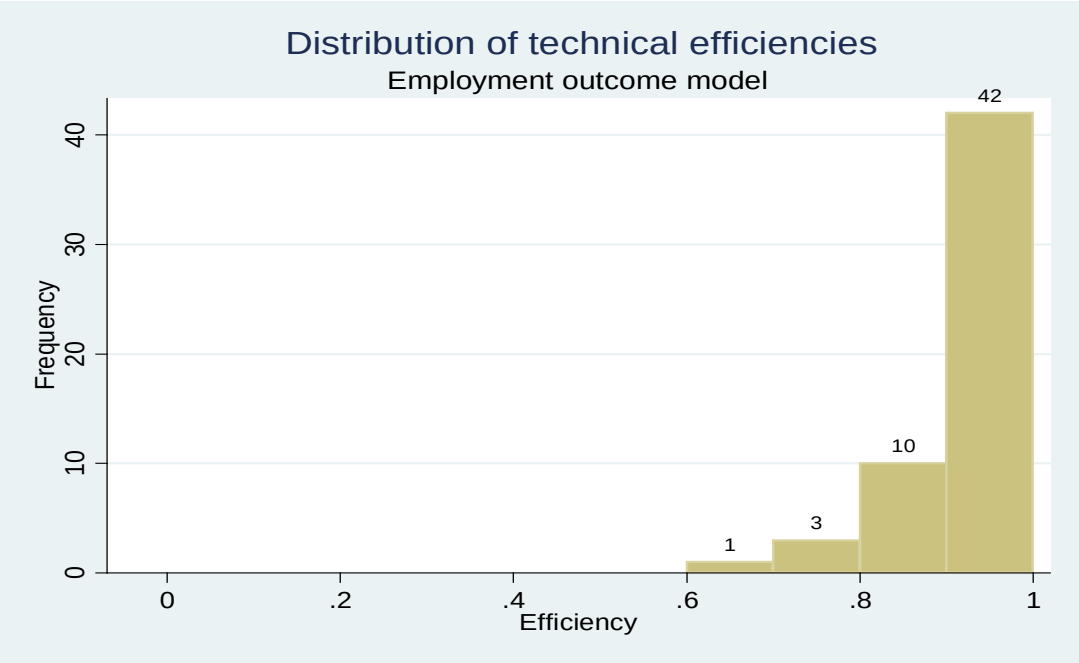

Figure 3.3 Distribution of SFA efficiencies: Employment outcome model

Although the two different efficiencies calculated were conceptually quite different, it was of interest whether institutes that scored highly in terms of one type of efficiency also did so on the other. We calculated the Pearson product moment correlation between both efficiencies as -0.099 ($p = 0.469$). It was thus apparent that the efficiencies for 'teaching hours' and 'employment outcomes' were statistically unrelated. (For a more detailed comparison between teaching hours and employment outcome efficiency, see section 3.1.10.)

Economies of scale effects can always be suspected where different units produce variable quantities of similar goods. In the context of the cost of teaching at a TAFE institute, this means that increasing 'hours taught' costs decrease on a per hour basis as operational fixed costs can be shared over more hours. In the higher education sector such economies of scale have been well documented (see, for instance, Hashimoto & Cohn, 1997; Laband & Lentz, 2003), albeit mostly in the university context. In the Australian TAFE sector, one could reasonably expect that larger institutes would exhibit higher

efficiency. We were therefore interested in patterns of efficiency with respect to institute size. Figure 3.4 displays the institutional efficiency in respect to institute size, as measured by teaching hours.

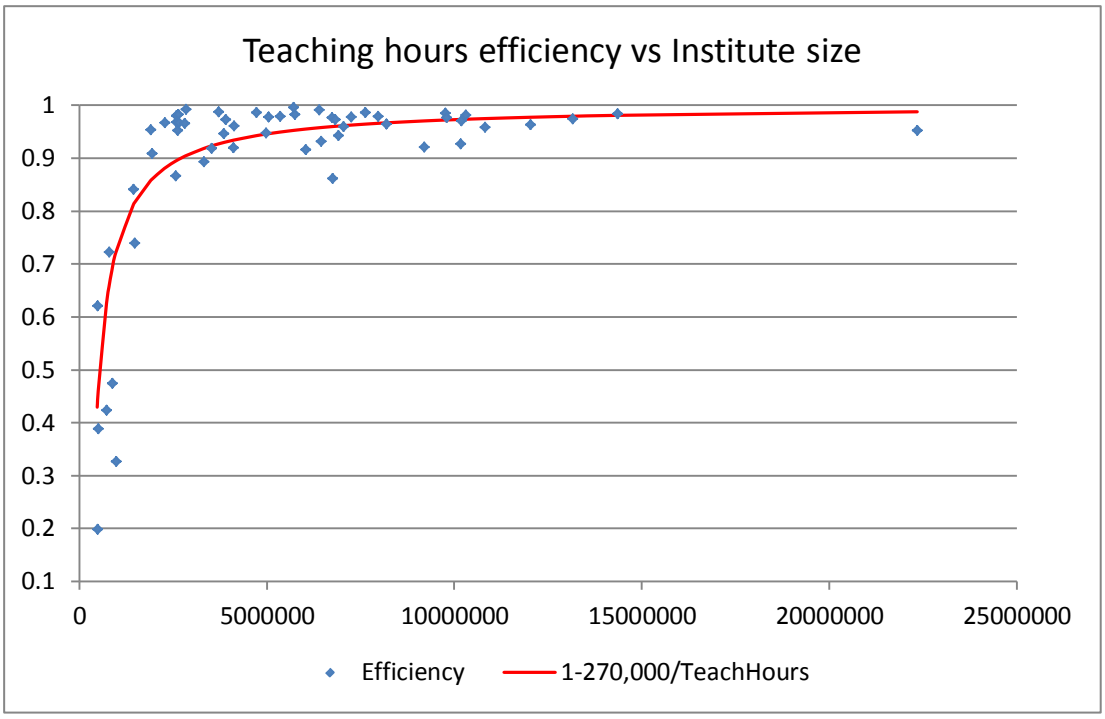

Figure 3.4 SFA 'Teaching hours' model efficiencies as a function of institute size

In this graph blue dots identify individual institutes and their location indicates the relationship between efficiency and size. As was hypothesised, smaller institutes appeared to exhibit significantly lower efficiencies than larger institutes. This graph should be of interest to regulators and policymakers, as it shows the dramatic change in efficiency over only a small portion of size increase on the far left of the chart. We fitted a curve over the data in order to be able to mathematically define the point at which further increases in size ceased to translate into significant gains in efficiency. Practically, this point should define the minimum size for a TAFE to operate efficiently. The curve fitted defines the relationship efficiency as a function of size as

$$E = 1 - \frac{2.7*10^5}{s} \qquad (23)$$

where S indicated size as measured by teaching hours. The resulting function explained about 88% of the variance in efficiency and was thus a reasonable representation of the data. We then defined the turning point of this function as the point where the strong increase in efficiency in respect to teaching hours diminishes. The derivative of (23) yields

$$\frac{dE}{ds} = \frac{2.7*10^5}{s^2} \qquad (24)$$

Solving (24) for a slope of 1 and accounting for the different scale of y and x axis yields

$$T = \sqrt{2.7 * 10^5 * TH_{max}} = 2.4 * 10^6 \qquad (25)$$

where TH_{max} represented the teaching hours of the largest institution. It can thus be stated that, based on the above derivation, when institutional size was equal to or greater than about 2.4 million teaching hours, size was no longer an impediment to efficiency. Alternatively it can be concluded that, in order to be efficient in transforming financial resources to units taught, TAFE institutes should be of a size that corresponds to at least 2.4 million teaching hours.

We were naturally interested if a similar pattern emerged when 'employment outcome' efficiency was mapped to TAFE size. The graph can be seen in figure 3.5.

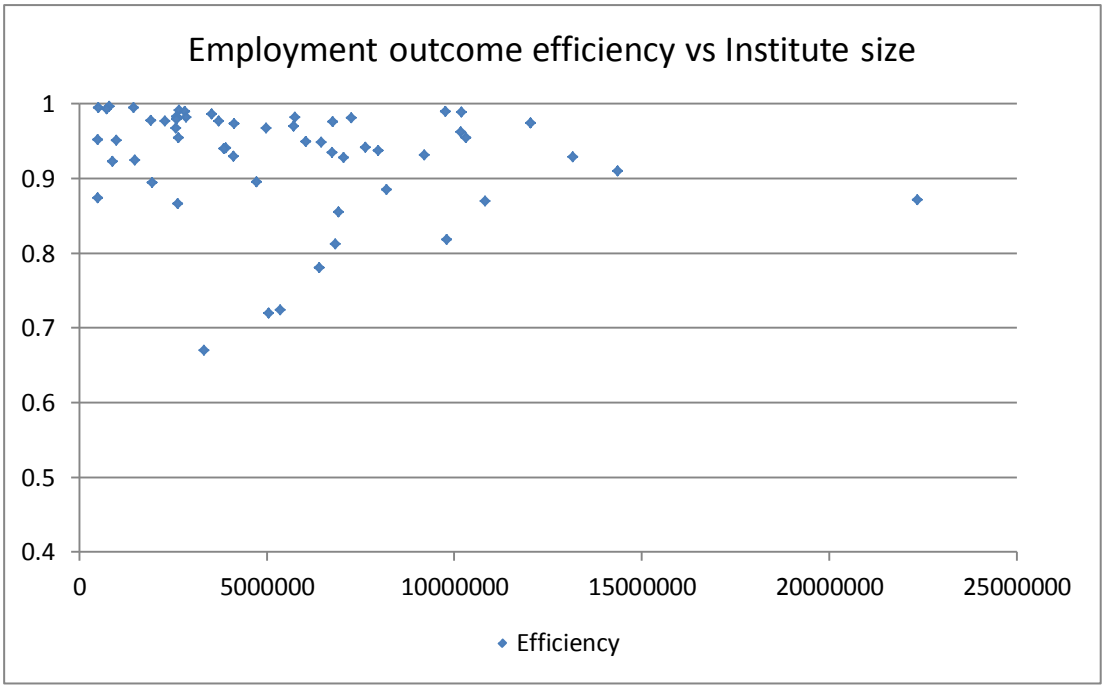

Figure 3.5 SFA 'Employment outcome' model efficiencies as a function of institute size

It was clear that there is no apparent pattern as in the 'teaching hours' efficiency graph. A linear regression yielded a coefficient with a non-significant negative slope. Considering the three DMUs below the 0.75 efficiency mark as outliers and removing them from the regression model produced a significant negative slope coefficient. However, it was also clear that the smaller institutes that displayed relatively low efficiencies in the 'teaching hours' model scored very well on the 'employment outcome' model. This demonstrated that care should be taken when interpreting these scores by themselves and that a multi-dimensional approach should be taken when making policy decisions based on indicators such as the one developed in this research.

3.1.8 DEA 'teaching hours' model

One of the aims of this study was to perform a comparative analysis of the results of SFA and DEA when applied to identical data. Both methods have a number of strengths and weaknesses and therefore each method could be deployed depending on the specific circumstances that applied to the respective topic under investigation. The main advantage of DEA was that no specific functional form needed to be specified for the calculation of

efficiencies and that DEA was capable of dealing easily with multiple input and output scenarios. On the other hand, DEA was sensitive to outliers, did not provide an easily interpretable model and was not able to exclude random error from the calculation of efficiencies. Given these advantages and constraints it was of interest to see how DEA performed with respect to the SFA presented in the previous section.

The topic of comparison between parametric and non-parametric production frontier comparison has been the subject of prior research in a variety of contexts. Becalli et al. (2006) found significant differences between SFA and DEA results when analysing the European banking sector. Whiteman (1999) applied SFA and DEA analysis to multiple output models in the utility and telecommunication sector. He found differing results by the two methods depending on which industry was being investigated. He concluded that DEA, despite its inability to exclude stochastic error, was the preferred technique, specifically in multiple output scenarios. Theodorides and Psychoudakis (2008) measured the efficiency of Greek dairy farms using SFA and DEA methods. They found relatively high correlations between DEA and SFA results, although they noted that the association between individual DEA results (for example, between constant returns to scale (CRS) and variable returns to scale (VRS)) was stronger than the association between DEA and SFA results. Other authors have focussed on the comparison of efficiency means resulting from both methods. Reinhard (1999) reported significant mean differences between efficiencies of Dutch dairy farms derived by SFA and CRS DEA. In that study SFA estimates of efficiency were higher than DEA estimates. Conversely, Kalaitzandonakis and Dunn (1995) reported notably higher efficiency means calculated by CRS DEA than estimated by SFA. It is worth noting that, due to the relative nature of calculated efficiency scores by either method (for example, both methods establish efficiency scores that rank institutions relative to one another), it was more instructive to compare parametric and

non-parametric efficiency scores via correlation analysis rather than differences in mean efficiency scores.

As there was no specification of a functional form necessary under DEA, the calculation of DEA efficiency scores for our 'teaching hours' model was relatively straightforward. The only decision that needed to be made was whether an input- or output-oriented DEA model would be more suitable. An input-oriented model essentially determines by how much the quantity of inputs can be reduced without changing the output produced. Conversely, an output-oriented model asks to what degree outputs can be increased while holding inputs constant. The DEA results of both orientations are identical for constant returns to scale analysis. However, they do differ under the variable returns to scale scenario. In our study we decided to employ an input oriented DEA, as institutions have more control over the inputs, for example, financial and administrative resources, rather than the number of students that are being taught at an institution.

We employed the data as described in the SFA section. We used the DEA extension in the STATA statistical software package as developed by Lee and Ji (2009) to derive efficiency scores. The results can be found in table 3.8.

Table 3.8 Comparison of DEA and SFA derived 'Teaching hours' model efficiency scores

Institute	CRS	VRS	SCALE	RTS	SFA
1	0.852	0.895	0.953	decreasing	0.984
4	0.780	0.803	0.971	decreasing	0.977
5	1.000	1.000	1.000	constant	0.973
7	0.708	0.715	0.990	increasing	0.932
10	0.949	1.000	0.949	decreasing	0.953
11	0.730	0.734	0.995	increasing	0.943
13	0.937	0.944	0.993	increasing	0.971
14	0.554	0.644	0.861	increasing	0.953
15	0.785	0.803	0.978	increasing	0.978
16	0.712	0.826	0.862	increasing	0.966
17	0.619	0.824	0.752	increasing	0.953
18	0.990	0.992	0.998	increasing	0.986
19	0.764	0.814	0.938	increasing	0.960
20	0.927	0.928	0.999	increasing	0.963
22	0.648	0.666	0.974	increasing	0.862
23	0.839	0.845	0.992	increasing	0.921
24	0.710	0.719	0.987	increasing	0.964
25	0.602	0.715	0.841	increasing	0.968
26	0.613	0.777	0.789	increasing	0.908
27	0.852	0.856	0.995	increasing	0.985
28	0.886	0.961	0.922	increasing	0.973
29	1.000	1.000	1.000	constant	0.959
30	0.913	1.000	0.913	increasing	0.987
31	0.710	0.837	0.848	increasing	0.967
32	0.613	0.722	0.849	increasing	0.866
33	0.694	0.799	0.868	increasing	0.982
34	0.812	0.844	0.962	increasing	0.996
35	0.783	0.835	0.938	increasing	0.920
36	0.713	0.732	0.975	increasing	0.986
37	0.916	0.959	0.955	increasing	0.979
38	0.887	0.914	0.970	increasing	0.991
40	0.389	1.000	0.389	increasing	0.621
43	0.911	0.932	0.977	increasing	0.960
44	0.754	0.805	0.936	increasing	0.946
45	0.682	0.749	0.911	increasing	0.893
46	0.641	0.743	0.862	increasing	0.980
47	0.858	0.890	0.964	increasing	0.916
48	0.832	0.838	0.994	increasing	0.927
49	0.676	0.773	0.875	increasing	0.992
50	0.829	0.833	0.995	increasing	0.972
51	0.862	0.910	0.948	decreasing	0.981
52	0.466	0.678	0.688	increasing	0.739
53	0.527	0.789	0.668	increasing	0.840
55	0.591	0.735	0.805	increasing	0.967
56	0.344	0.652	0.529	increasing	0.474
57	0.393	0.774	0.508	increasing	0.723
58	0.255	0.495	0.515	increasing	0.327

Continued next page

Institute	CRS	VRS	SCALE	RTS	SFA
60	0.235	0.612	0.384	increasing	0.389
64	0.671	0.687	0.977	increasing	0.948
65	0.794	0.807	0.984	increasing	0.977
66	0.824	0.837	0.985	increasing	0.979
70	0.668	0.725	0.921	increasing	0.918
71	1.000	1.000	1.000	increasing	0.978
74	0.161	1.000	0.161	increasing	0.198
77	0.588	0.611	0.963	increasing	0.983
110	0.329	1.000	0.329	increasing	0.423
Mean	0.710	0.821	0.862		0.888
Std dev	0.203	0.121	0.196		0.182

The constant return to scale (CRS), variable returns to scale (VRS), and scale efficiency (SCALE) columns contain the respective calculated efficiencies. The returns to scale (RTS) column denotes decreasing (DRS), constant (CRS), or increasing returns (IRS) to scale characteristics. The SFA column displays the efficiency estimates from the parametric frontier analysis. The calculated means for all four types of efficiency are 0.71, 0.82, 0.86, and 0.89 for CRS, VRS, SCALE and SFA respectively. VRS efficiency has the lowest variability whereas CRS has the highest. The distributions of the three DEA efficiencies can be found in figures 3.6, 3.7 and 3.8, and for comparison, SFA efficiencies are displayed in figure 3.9.

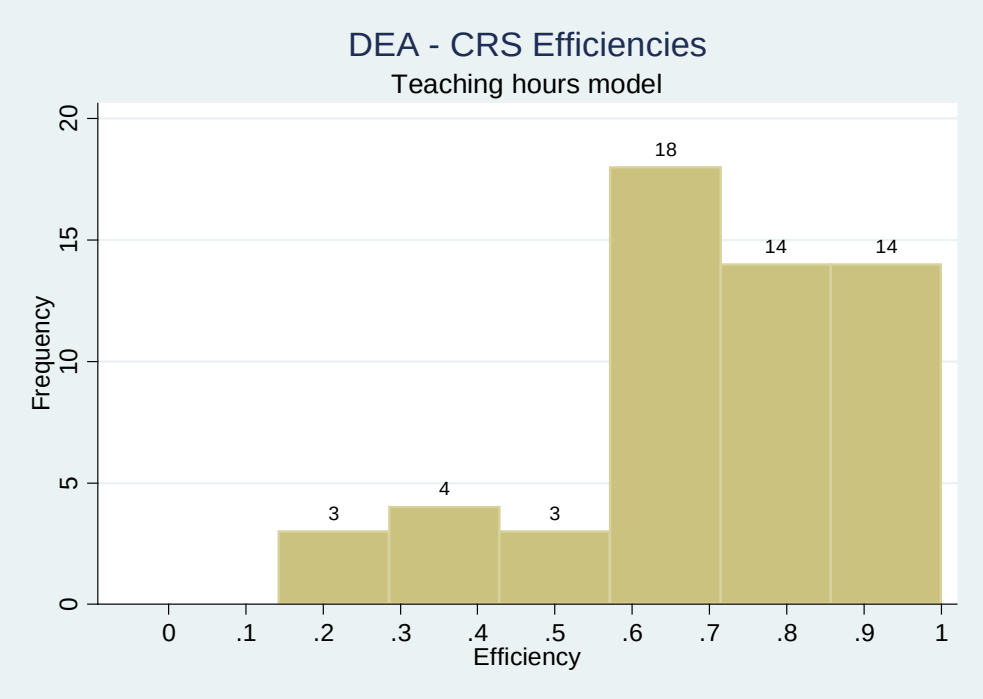

Figure 3.6 Distribution CRS efficiencies: Teaching hours model

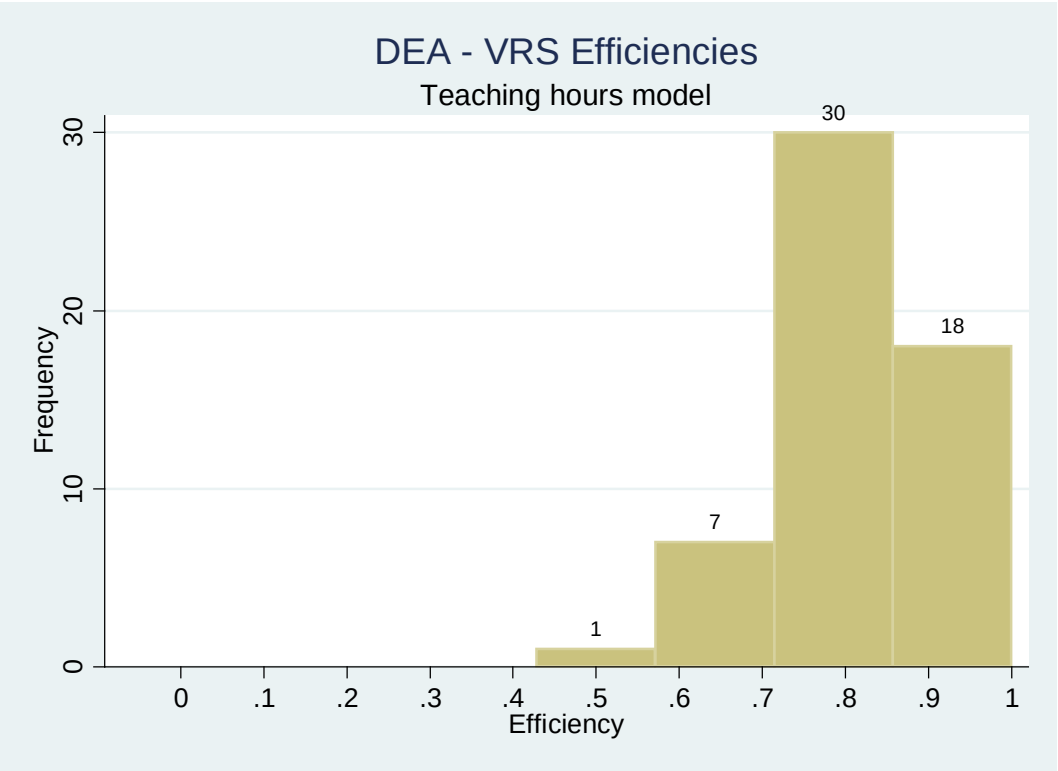

Figure 3.7 Distribution VRS efficiencies: Teaching hours model

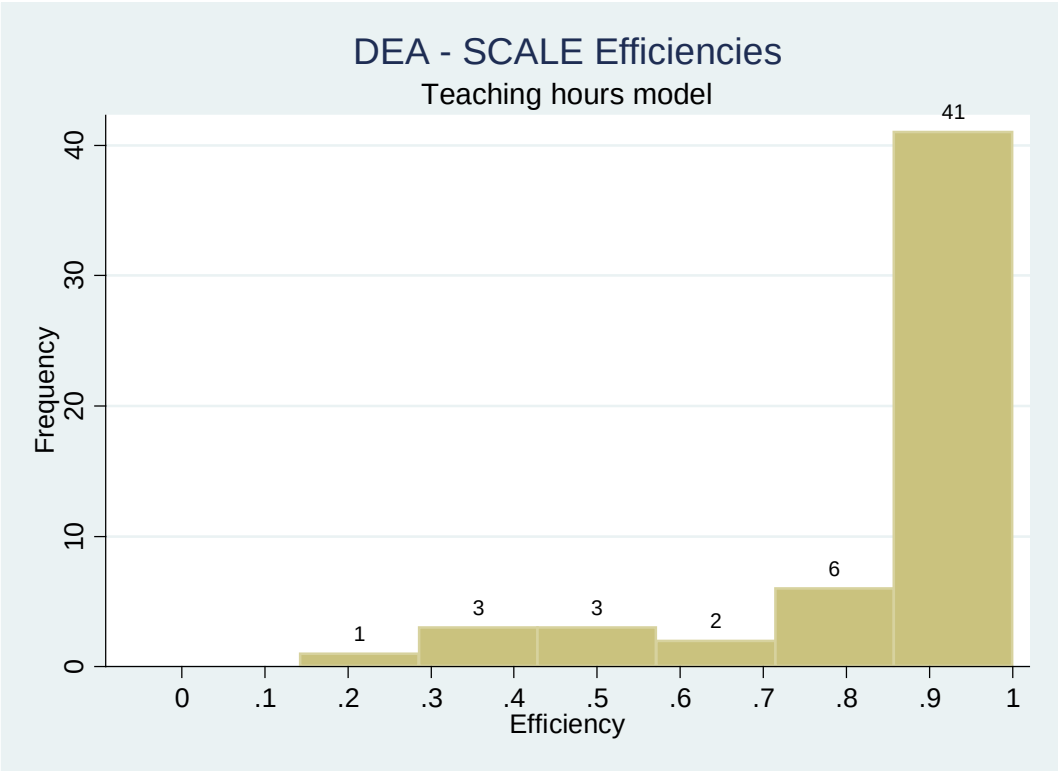

Figure 3.8 Distribution SCALE efficiencies: Teaching hours model

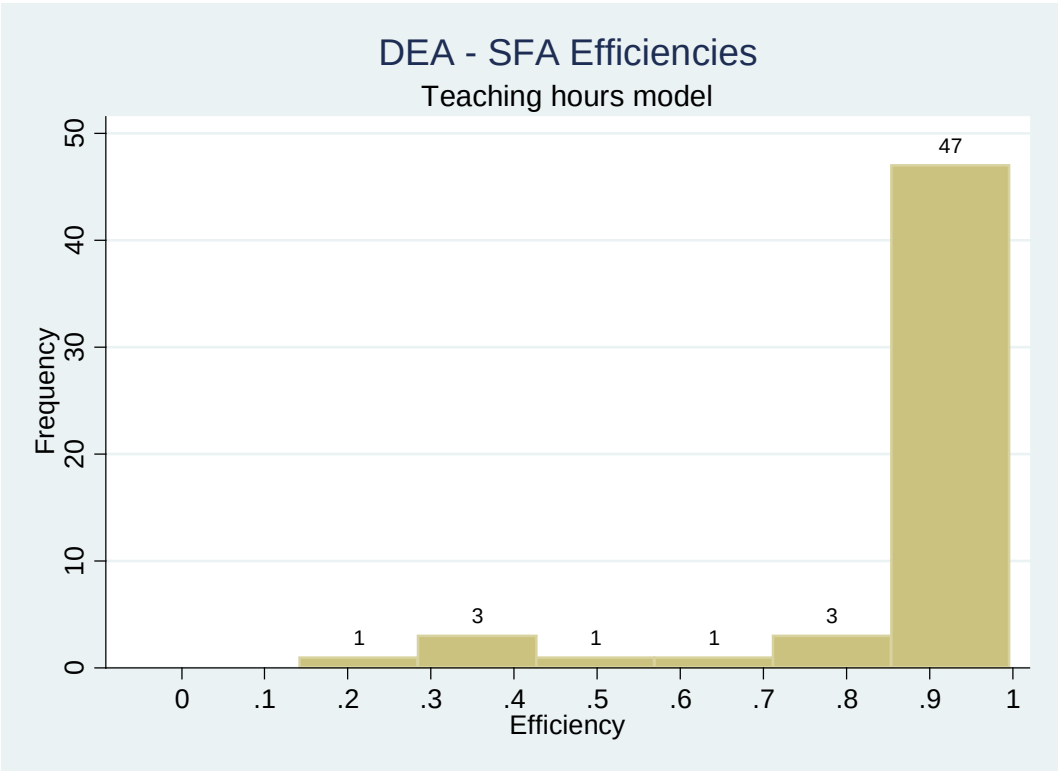

Figure 3.9 Distribution SFA efficiencies: Teaching hours model

Most institutes displayed increasing returns to scale technology, signifying that increasing inputs disproportionally increased 'teaching hours' output. Our main interest was in examining the relationship between the three DEA and the SFA efficiencies. We computed a pair-wise correlation matrix with results shown in table 3.9.

Table 3.9 Correlation matrix of various 'Teaching hours' efficiencies

	CRS	VRS	SCALE	SFA
CRS	1			
VRS	0.5387	1		
	<.0001			
SCALE	0.8815	0.1067	1	
	<.0001	0.4336		
SFA	0.8235	0.1775	0.9099	1
	<.0001	0.1906	<.0001	

The strongest correlation was between SFA and SCALE efficiency (0.91; $p<0.001$), followed closely by CRS and SCALE efficiency (0.88; $p<0.001$) and CRS and SFA (0.82; $p<0.001$). VRS efficiency was only moderately correlated with CRS efficiency (0.54; $p<0.001$) and not correlated with SCALE efficiency (0.11; $p = 0.434$) and SFA efficiency (0.17; $p = 0.191$). The more robust correlation between SFA and CRS than between SFA and VRS has been observed frequently (see Oren & Almedar, 2006; Theodorides & Psychoudakis, 2008; Prochazkova, 2011; Fatulescu, 2013; although Önder et al., 2013 noted a stronger association between SFA and VRS than between SFA and CRS). The strength of the correlation between SFA and CRS itself (0.82) gives reason to be confident about the robustness of the patterns of institutional efficiency estimates, given that they were derived via two different methods using differing assumptions. In the 'teaching hours' SFA model we calculated a γ estimate of 0.9. This high value pointed to the existence of a deterministic efficiency frontier. Given the relative consistency of stronger associations between SFA and CRS in this study (and also seen in other published empirical studies), and the large γ value obtained from our SFA teaching hours model, it

may be concluded that the production frontiers estimated by both methods have a similar shape.

Since we have established a strong association between SFA and DEA CRS results we hypothesized that our inefficiency predictors from the SFA model also predicted inefficiencies in the DEA CRS model. To test this hypothesis we employed the CRS efficiency scores created by the DEA model as a dependent variable and the inefficiency predictors from the SFA model as an independent variable in a Tobit regression model, as suggested by Bjurek et al.(1992). Tobit models are applied when a dependent variable is censored to one or both sides and thus an OLS model estimates would exhibit truncation bias. Our DEA scores were by definition censored on 1, so that right censoring applied. The estimation results of the Tobit model can be seen in table 3.10.

Table 3.10 Predictors of DEA CRS and SFA teaching hours efficiency

	Tobit model – 'Teaching hours' CRS efficiency predictors				Inefficiency SFA	
Variable	Estimate	Std error	t	P>|t|	Estimate	P>|t|
Constant	1.506	0.325	4.640	<.001	-17.631	0.001
English second language	-0.002	0.002	-0.760	0.450	0.129	0.027
Students with disability	0.009	0.007	1.290	0.203	0.053	0.726
Remoteness (ARIA)	-0.158	0.021	-7.370	<.001	2.708	<.001
Student age	-0.014	0.009	-1.460	0.150	-0.074	0.768
Proportion of males	-0.001	0.002	-0.640	0.525	0.112	0.048
Sigma	0.126	0.012	10.156	<.001		
	Chi-sq = 54.67 P >|chisq| <0.001					

The only significant predictor of efficiency was the remoteness variable and indicated decreasing efficiency as remoteness increased. This finding replicated the results from the SFA inefficiency model. In the SFA model increasing English as a second language and the proportion of males were also significant predictors of inefficiency. While these two variables were not significant in the DEA Tobit model, their signs indicated a relationship in the same direction (in table 3.10 the signs are reversed, due to the DEA CRS Tobit predicting efficiency and the SFA predicting inefficiency). Overall we can thus

assert that the predictors for SFA and CRS DEA scores indicated similar associations with efficiency.

3.1.9 DEA employment outcome model

For the examination of the employment model using the DEA approach we utilised the same data as in the SFA employment model. A minor modification was necessitated by the inability of DEA to deal with categorical variables. We thus recoded the institutional size indicator and treated it as an interval variable (1 indicating very small to 5 indicating very large) as proposed by Banker and Morey (1986). The DEA for the employment outcome results can be found in table 3.11.

Table 3.11 Comparison of DEA and SFA derived 'Employment outcome' model efficiency scores

Institute	CRS	VRS	SCALE	RTS	SFA
1	0.386	0.810	0.477	increasing	y
4	0.196	0.850	0.230	increasing	0.820
5	0.450	0.838	0.538	increasing	0.927
7	0.611	0.870	0.702	increasing	0.950
10	0.317	0.771	0.411	increasing	0.870
11	0.412	0.927	0.444	increasing	0.860
13	0.820	0.965	0.849	increasing	0.989
14	0.581	0.838	0.693	increasing	0.859
15	0.803	0.879	0.913	increasing	0.982
16	0.924	1.000	0.924	increasing	0.991
17	0.739	1.000	0.739	increasing	0.981
18	0.670	1.000	0.670	increasing	0.944
19	1.000	1.000	1.000	constant	0.973
20	0.477	0.857	0.556	increasing	0.973
22	0.588	0.865	0.680	increasing	0.976
23	0.392	0.847	0.462	increasing	0.924
24	0.297	0.655	0.453	increasing	0.878
25	0.700	0.958	0.731	increasing	0.980
26	0.526	0.930	0.566	increasing	0.896
27	0.600	0.901	0.667	increasing	0.990
28	0.796	0.952	0.836	increasing	0.939
29	0.284	0.885	0.321	increasing	0.871
30	1.000	1.000	1.000	constant	0.978
31	1.000	1.000	1.000	increasing	0.992
32	0.848	0.915	0.928	increasing	0.968
33	0.769	0.968	0.795	increasing	0.956
34	0.574	0.870	0.660	increasing	0.973
35	0.766	0.907	0.845	increasing	0.929

Continued next page

Institute	CRS	VRS	SCALE	RTS	SFA
36	0.589	0.809	0.727	increasing	0.891
37	0.226	0.830	0.272	increasing	0.719
38	0.219	0.732	0.299	increasing	0.780
40	0.730	1.000	0.730	increasing	0.946
43	0.508	0.857	0.593	increasing	0.926
44	0.810	0.941	0.860	increasing	0.941
45	0.145	0.864	0.168	increasing	0.669
46	1.000	1.000	1.000	constant	0.983
47	0.600	0.864	0.695	increasing	0.955
48	0.573	0.881	0.650	increasing	0.963
49	0.823	0.923	0.892	increasing	0.983
50	0.278	0.867	0.321	increasing	0.819
51	0.513	0.870	0.590	increasing	0.954
52	0.697	1.000	0.697	increasing	0.938
53	1.000	1.000	1.000	constant	0.995
55	0.926	0.967	0.958	increasing	0.978
56	0.819	1.000	0.819	increasing	0.932
57	1.000	1.000	1.000	constant	0.997
58	0.986	1.000	0.986	increasing	0.953
60	1.000	1.000	1.000	constant	0.995
64	0.699	0.876	0.799	increasing	0.969
65	0.547	0.889	0.615	increasing	0.940
66	0.533	1.000	0.533	increasing	0.947
70	1.000	1.000	1.000	constant	0.986
71	0.359	1.000	0.359	increasing	0.724
74	1.000	1.000	1.000	constant	0.885
77	0.847	0.939	0.902	increasing	0.983
110	1.000	1.000	1.000	constant	0.994
Mean	0.660	0.917	0.706		0.929
Std dev	0.254	0.080	0.237		0.074

It was clear that constant-returns-to-scale and scale efficiencies were significantly lower than variable-returns-to-scale efficiencies, while exhibiting much larger variability. Figures 3.10, 3.11, 3.12 and 3.13 show the distributions for the DEA and SFA results for the employment model.

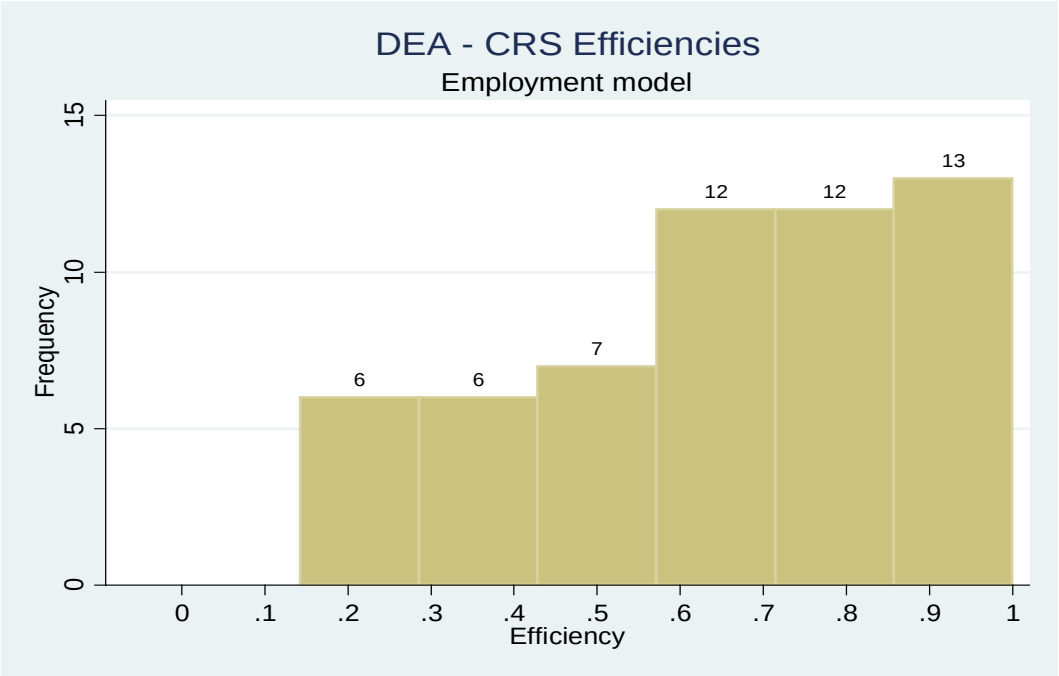

Figure 3.10 Distribution of CRS efficiencies: Employment outcome model

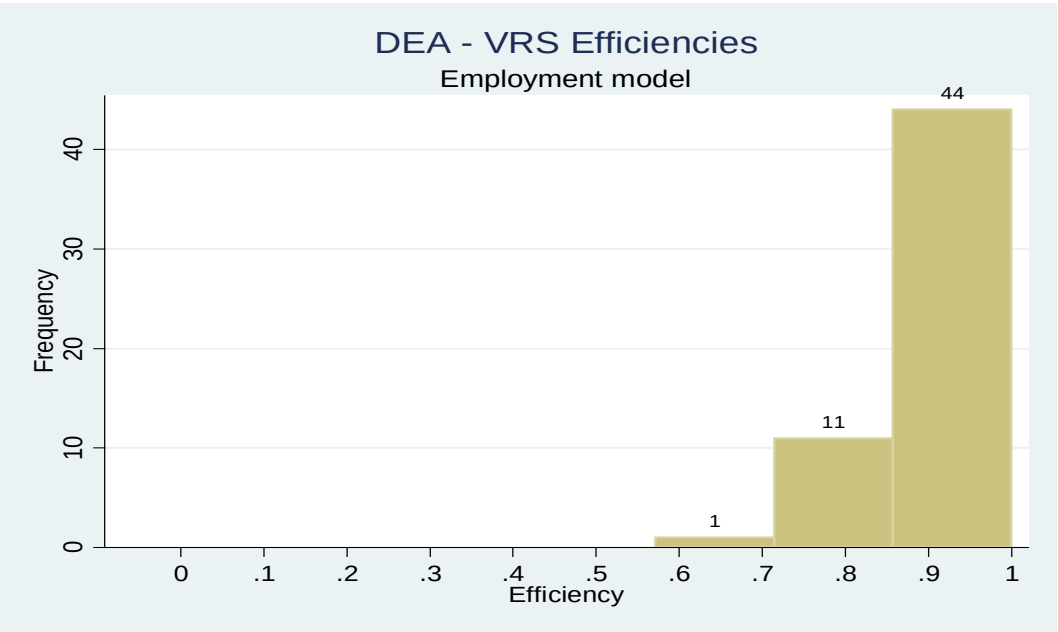

Figure 3.11 Distribution of VRS efficiencies: Employment outcome model

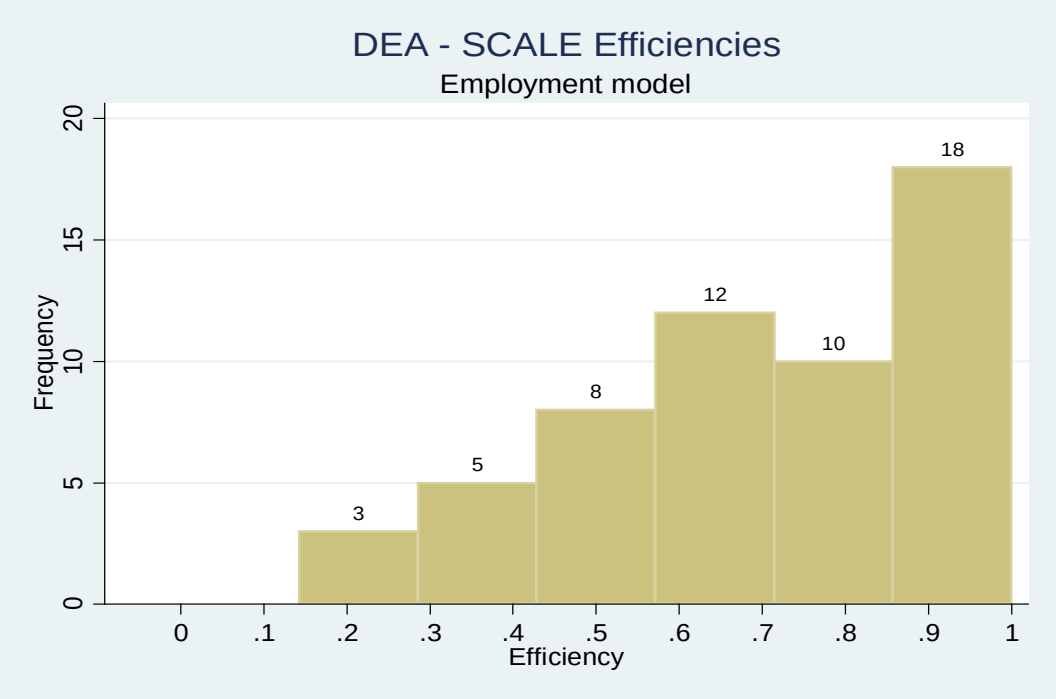

Figure 3.12 Distribution of SCALE efficiencies: Employment outcome model

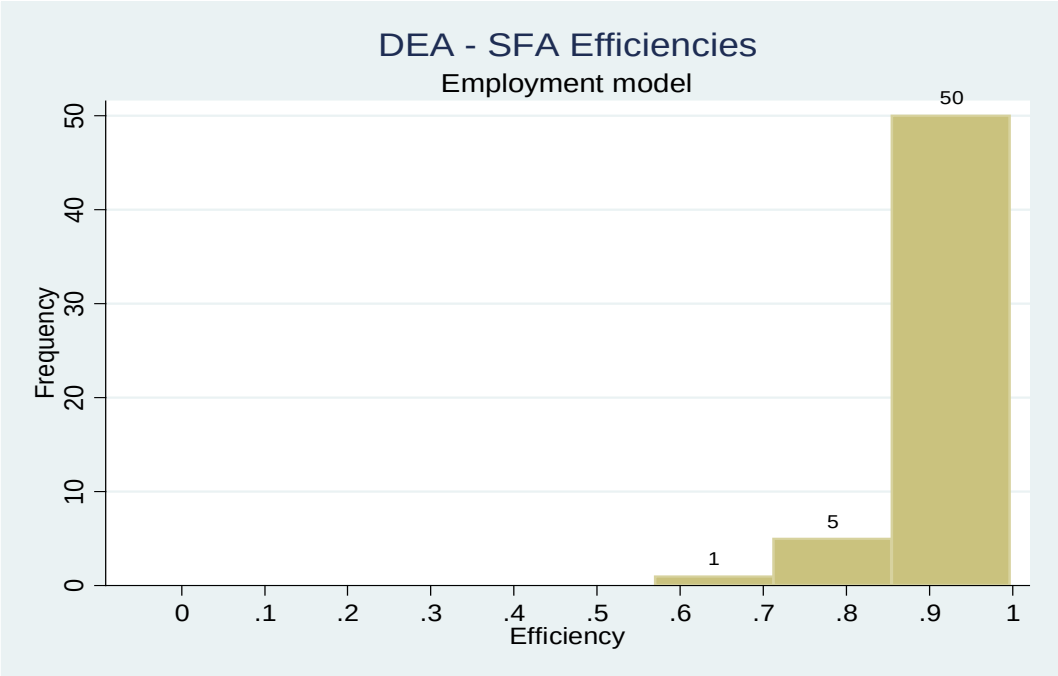

Figure 3.13 Distribution of SFA efficiencies: Employment outcome model

While means and distributions differed significantly between individual DEA and SFA efficiencies, we were predominantly interested in the correlations between the efficiency scores produced by both methods. Pairwise

correlations between the DEA employment model efficiency scores and respective SFA scores are displayed in table 3.12.

Table 3.12 Correlations between employment model efficiencies

	CRS	VRS	SCALE	SFA
CRS	1			
VRS	0.7466	1		
	<.0001			
SCALE	0.9876	0.6451	1	
	<.0001	<.0001		
SFA	0.7514	0.4386	0.7858	1
	<.0001	0.0007	<.0001	

All efficiency types correlated significantly to varying degrees. As in the 'teaching hours' models, correlations between SFA and DEA CRS were much higher than between DEA VRS and SFA. Looking at the Tobit model predictors (table 3.13)[13] we noted that there were more significant predictors of efficiency in the DEA CRS model than in the SFA model. The most dominant relationship in DEA and SFA was between employment outcome efficiency and remoteness. This relationship was negative indicating that employment outcome efficiency decreased as remoteness increased. The proportion of students with a disability reduced employment outcome efficiency significantly under DEA and SFA methodology. All other significant predictors for efficiencies were only found in the DEA model.

[13] We attached the results of the SFA inefficiency predictors from the SFA (table 3.5) for comparison. Note the signs of the estimates are reversed, as the SFA model predicts inefficiency rather than efficiency.

Table 3.13 Employment model efficiency predictors

	Tobit model – CRS efficiency predictors				Inefficiency SFA	
Variable	Estimate	Std error	t	P>|t|	t	P>|t|
Constant	0.745	0.678	1.100	0.278	-1.610	0.871
English second language	-0.009	0.004	-2.480	0.017	0.078	0.112
Students with disability	-0.026	0.012	-2.250	0.030	0.416	0.061
Remoteness (ARIA)	-0.227	0.046	-4.920	<.001	2.233	0.004
Student age	0.018	0.016	1.110	0.275	-0.493	0.076
Proportion of males	-0.001	0.003	-0.320	0.751	-0.003	0.944
Funding per hour	0.014	0.004	3.230	0.002	-0.044	0.331
Completion rate	-0.006	0.002	-2.470	0.018	0.048	0.495
Cert III or higher	-0.006	0.004	-1.650	0.107	0.041	0.642
Graduates	-0.003	0.002	-1.380	0.174	0.017	0.684
Load pass rate	0.008	0.005	1.580	0.121	-0.025	0.735
Very large	-	-	-	-	-	-
Large	-0.021	0.072	-0.290	0.775	1.333	0.272
Medium	0.179	0.081	2.210	0.032	1.320	0.286
Small	0.297	0.084	3.520	0.001	-0.689	0.685
Very small	0.667	0.167	3.980	<.001	-4.861	0.275
Sigma	0.152	0.016	9.423	<.001		
	Chi-sq = 73.02		p >|chisq| <0.001			

Higher proportions of students with English as a second language and larger completion rates were both associated with decreased employment outcome efficiency. While this appeared to be a logical result for the proportion of students with English as a second language who may experience difficulty in obtaining employment based on language issues, it seemed counterintuitive that a higher completion rate should lead to lower employment outcome efficiency. There may be several reasons for this. Mark and Karmel (2010) found that one of the characteristics of the vocational education sector was that many students wished to learn specific skills and had no desire to complete a full qualification. Therefore, large numbers of students may leave their training prematurely once they have the skills necessary to obtain employment. Another reason for the decreasing employment outcomes with increasing completion rates may be that there were significant numbers of students who entered the vocational education system on government-sponsored programs. The students in such programs do tend to have high completion rates while at the same time achieving worse employment

outcomes than non-sponsored students. An example of this was the 2009 Productivity Placements Program (NCVER, 2009) where this trend was particularly apparent. Finally, another possible explanation may be in the way completion rates were derived. As discussed in section 3.1.5, there was no universal agreement on the best way to determine completion rates. Therefore, completion rates may vary depending on what method was used to calculate them, which may lead to changing results when using them in the prediction of employment outcome efficiency.

Funding per teaching hour was related to an increase in employment outcome efficiency under the DEA CRS model while it was not significant in the SFA model. There was a significant trend of smaller institutes being more employment outcome efficient than larger institutes.

3.1.10 Relationship between 'Teaching hours' efficiency and 'Employment outcome' efficiency

Examining the SFA graphs of 'teaching hours' efficiency as a function of institute size (figure 3.3) and 'employment outcome' efficiency as a function institute size (figure 3.4) it was unlikely that there was a strong relationship between the two efficiencies. An analysis of correlations between the four different types of efficiency calculated in the study (table 3.14) showed negative correlations across the board. These correlations were significant in the case of CRS and SCALE efficiencies albeit fairly weak. SFA and VRS efficiencies of the 'teaching hours' model and 'employment outcome' model were not correlated. It can thus be concluded that institutes that were efficient in the conversion of financial and administrative resources into teaching output may not necessarily be so in the efficiency of achieving employment outcomes for their students. To further investigate a possible relationship between teaching hours efficiency and employment outcome efficiency we graphed the two measures in a scatterplot (figure 3.14).

Table 3.14 Correlations between teaching hours and employment outcome efficiencies

	SFA		CRS		VRS		SCALE	
	Teaching hours efficiency	Employm. outcome efficiency	Teaching hours efficiency	Employm. outcome efficiency	Teaching hours efficiency	Employm. outcome efficiency	Teaching hours efficiency	Employm. outcome efficiency
Teaching hours efficiency	1		1		1		1	
Employm. outcome efficiency	-0.1134 0.4052	1	-0.5671 <0.001	1	-0.0741 0.5871	1	-0.4872 <0.001	1

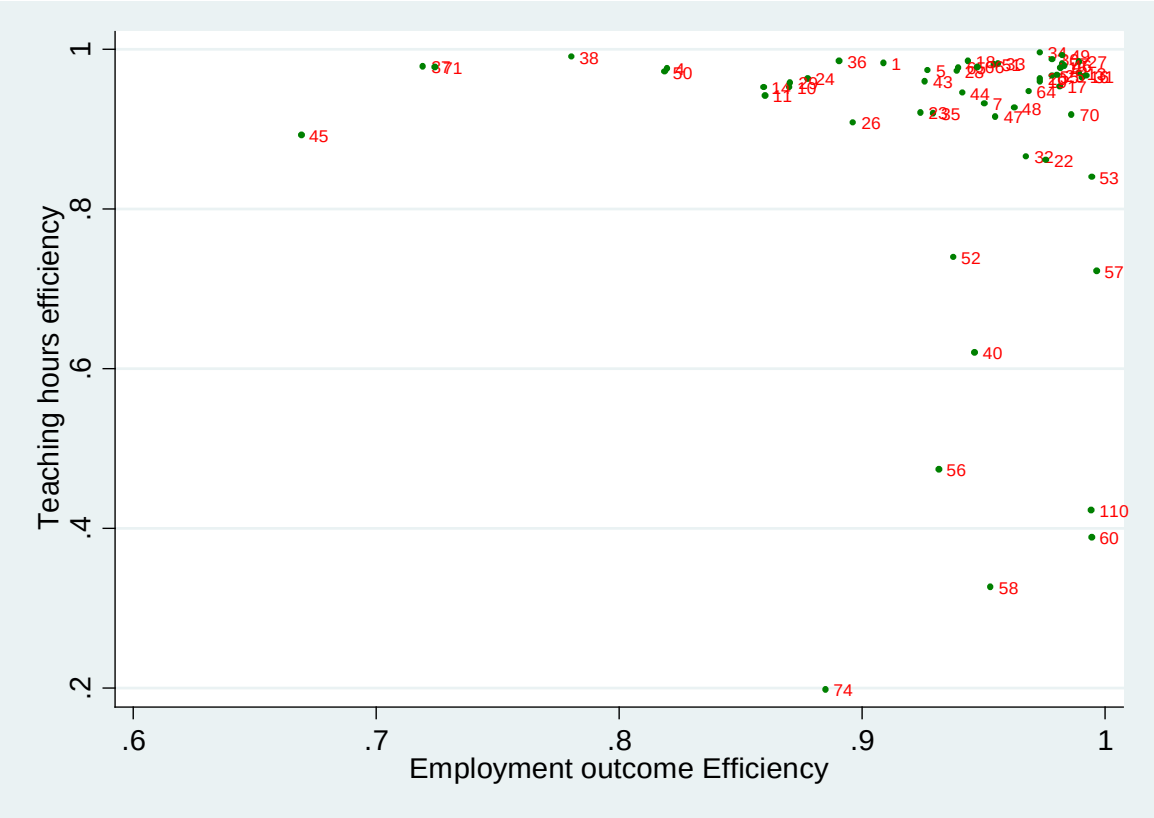

Figure 3.14 Location of institutes in teaching hours and employment outcome efficiency

An interesting pattern became evident from this graph. There appeared to be three major constellations. Some institutes scored relatively low on 'teaching hours' efficiency and high on employment outcome efficiency, whereas others attained a high teaching hours efficiency and low employment outcome efficiency and the remainder rated relatively high on both efficiencies. Interestingly, there were no institutions that displayed low scores on both types of efficiencies examined in this study. It was of interest to statistically separate these three possible combinations of teaching hours and employment

outcome efficiency (for example, high/high, high/low and low/high) and to evaluate the institutions that constituted the pattern in figure 3.14 with respect to possible demographic, educational and environmental variables as determinants of group membership thereof. We performed a partition cluster analysis, using the k-means method with three target clusters. This technique involved an iteration process in which each institute was initially randomly assigned to a cluster, and then subsequently was allocated to the cluster with the closest mean, as calculated using the Euclidean distance method. After this, new cluster means were determined and the process iteratively continued until no institute changed groups. The resulting clusters can be seen in table 3.15.

Table 3.15 Institute location

Location	Institute
Location 1	40, 56, 58,60, 74, 110
Location 2	4, 10, 11, 14, 24, 29, 37, 38, 45, 50, 71
Location 3	1, 5, 7, 13, 15, 16, 17, 18, 19, 20, 22, 23, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 43, 44, 46, 48, 49, 51, 52, 53, 55, 57, 64, 65, 66, 70, 77

The location allocation following from the clusters in table 3.15 can be seen in figure 3.15.

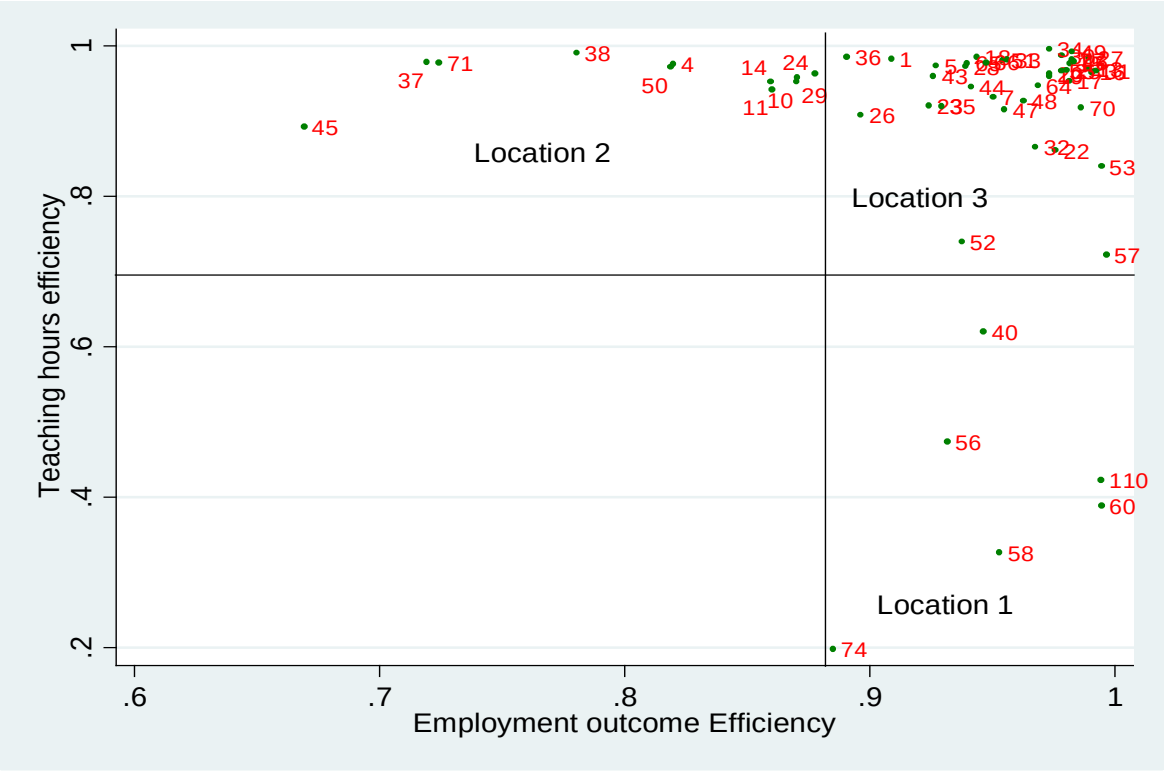

Figure 3.15 Location of institutes by group

We employed canonical discriminant analysis to examine the extent to which several covariates could be utilised to statistically differentiate between locations 1, 2 and 3. The covariates entered into the discriminant function were age, completion rate, load pass rate, disability (%), remoteness, graduates (%), age, male gender (%), satisfaction, salary, indigeneity (%), SES, certificate III or higher (%), English as a second language (%), Australian born (%), the percentage of apprentices and trainees and the size of the institution as measured by the number of student delivery hours. The essential statistics for the two resulting discriminant functions can be found in table 3.16.

Table 3.16 Canonical discriminant functions

Discriminant Function	Canonical Correlation	Eigenvalue	Cumulative Variance	Likelihood Ratio	F	Pr>|F|
1	0.864	2.937	0.787	0.141	3.937	<0.001
2	0.665	0.794	1.000	0.558	2.064	0.035

It can be seen that both discriminant functions were significant, but that the first discriminant function captured 80% of the variance. The discriminating

ability of the covariates was then be assessed by the evaluation of the standardised canonical discriminant function coefficients (table 3.17).

Table 3.17 Standardised canonical discriminant function coefficients

	Function 1	Function 2
Load pass rate	0.222	0.021
Completion rate	-0.297	-0.601
Disability %	0.477	-0.163
Remoteness	-0.927	-0.037
Graduates %	0.136	-0.070
Age	0.300	1.100
Male %	-0.350	-0.110
Satisfaction	0.113	-0.151
Salary	-0.268	0.001
Indigenous %	-1.117	-0.632
SES	0.007	0.629
Certificate III or higher %	0.141	-0.004
English as a second language %	0.591	0.313
Australian born %	1.043	0.470
Apprentices and trainees %	0.436	0.794
Institute size (in million delivery hours)	-0.177	-0.422

Generally, values close to zero indicated diminishing discriminating ability to separate the three locations. The percentage of disabled students, for instance, had thus a negligible contribution to the separability of the three efficiency locations. The discriminant function coefficients were graphed for easier interpretation (figure 3.16). Variables near the origin of this graph, such as load pass rate, certificate III or higher, student satisfaction, and percentage of graduates provided little discriminating ability. The location of the remaining variables signified their contribution to the discriminant function, with age, remoteness and percentage Indigenous and Australian-born students and apprentices and trainees having the strongest impact.

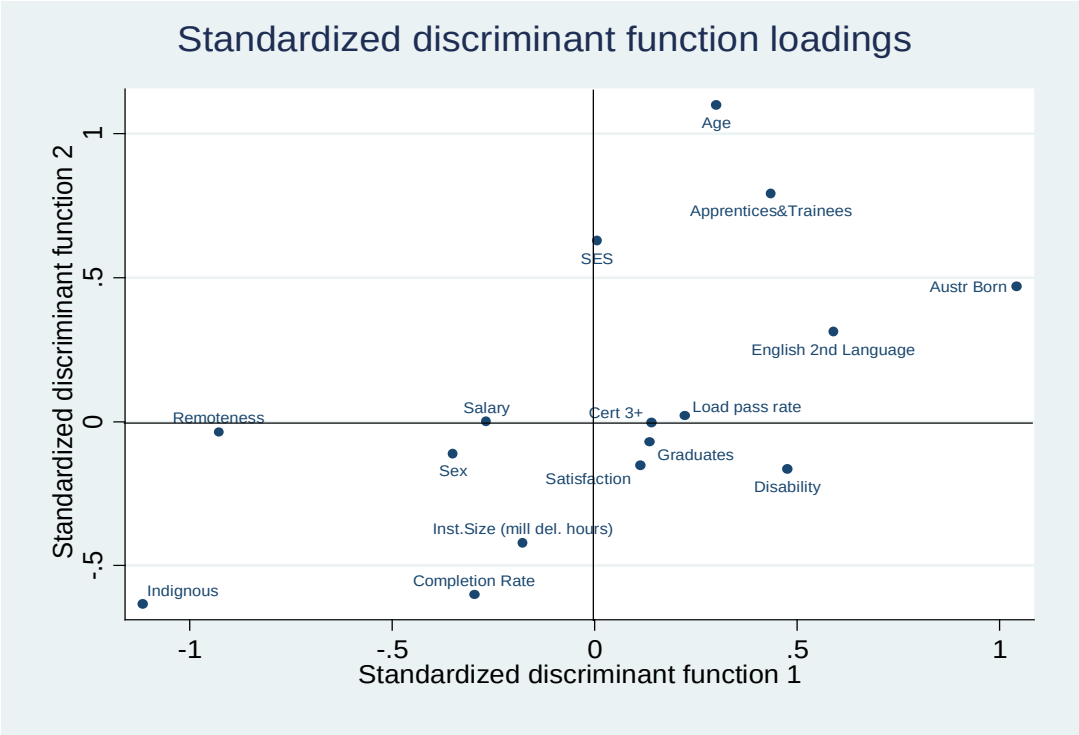

Figure 3.16 Standardised discriminant function loadings

Finally, we examined the confusion matrix (table 3.18) and the discriminant function plot (figure 3.17) to assess how well the covariates are able to separate the three efficiency locations.

Table 3.18 Confusion matrix

Location	**1**	**2**	**3**	**Total**
TRUE		Classified		
1	6	0	0	**6**
	100	0	0	**100**
2	0	8	3	**11**
	0	72.7	27.3	**100**
3	0	1	38	**39**
	0	2.56	97.4	**100**
Total	6	9	41	**56**
	10.7	16.1	73.2	**100**
Priors	0.11	0.20	0.70	**100**

Table 3.18 illustrates how many institutions were correctly classified into their location using the two significant discriminant functions. Overall 52 of the 56 institutes (92.9%) were accurately classified. Locations 2 and 3 appeared to have more misclassifications, implying that these two locations were harder to

separate. Examination of the discriminant function score plot (figure 3.17) confirmed that location 1 was fairly well separated from the others, while there was some notable overlap between locations 2 and 3.

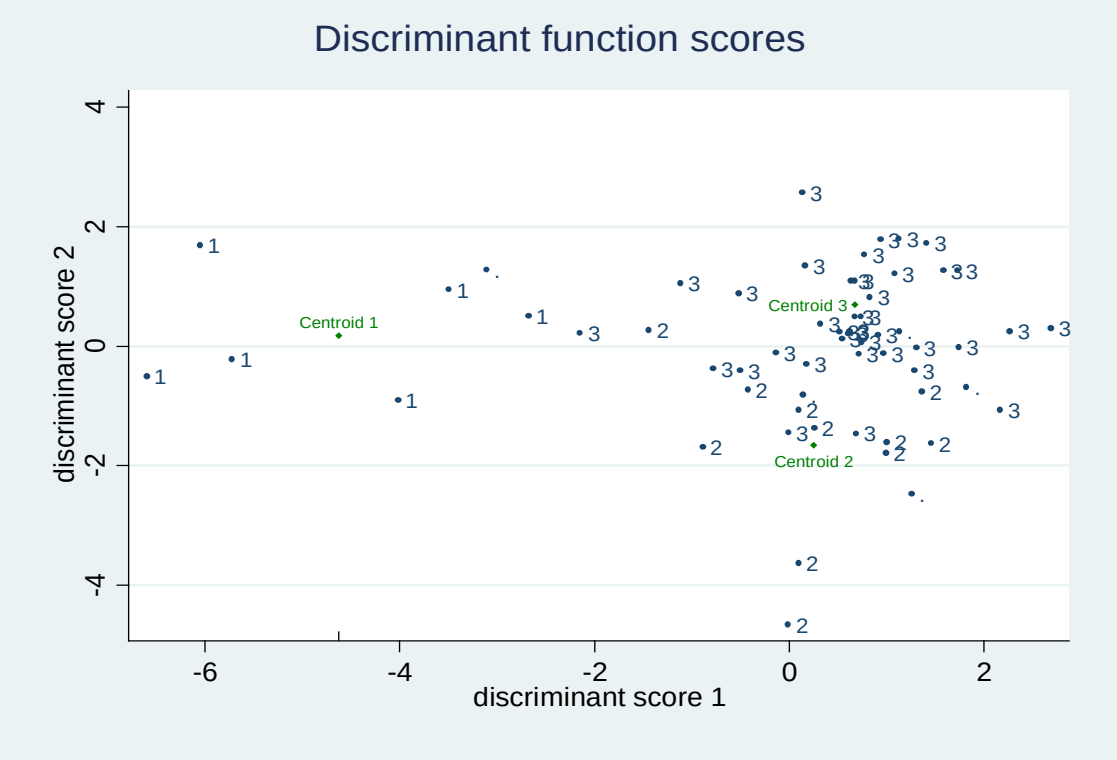

Figure 3.17 Discriminant function score plot

Finally, we calculated the means of the covariates of the canonical discriminant analysis and performed a one-way analysis of variance including a Bonferroni multiple comparison test. The results can be found in table 3.19.

Table 3.19 Location means and comparison tests

	Location means			Location differences P>|t|			P>|F|
	1	2	3	1v2	1v3	2v3	
Load pass rate	78.6	79.3	82.7	1.000	0.471	0.420	0.169
Completion rate	15.5	36.2	26.6	0.001	0.061	0.033	0.001
Disability %	7.9	10.5	9.3	0.231	0.801	0.660	0.195
Remoteness	4.0	1.8	1.9	<.001	<.001	1.000	<.001
Graduates %	27.2	49.6	37.7	0.004	0.218	0.031	0.004
Age	34.2	31.5	33.2	0.046	0.855	0.071	0.028
Male %	63.9	51.6	57.7	0.068	0.521	0.274	0.063
Satisfaction	4.3	4.2	4.2	0.036	0.188	0.481	0.041
Salary	68,814	53,225	55,990	<.001	<.001	0.442	<.001
Indigenous %	24.3	6.3	3.0	0.002	<.001	1.000	<.001
SES	2.4	2.9	3.0	0.592	0.134	1.000	0.124
Cert III or higher %	73.3	81.7	83.5	0.154	0.020	1.000	0.024
English second language %	14.2	20.0	15.6	0.732	1.000	0.567	0.359
Australian born %	84.3	77.5	79.7	0.538	0.878	1.000	0.402
Apprentices and trainees %	18.0	15.5	17.3	1.000	1.000	1.000	0.736
Inst. size (million del. hours)	0.7	8.0	5.6	0.001	0.013	0.210	0.002

The table confirmed that differences were more prominent between location 1 vs 2 and 3 rather than between locations 2 and 3. Completion rates stood out as being statistically different between all three locations, with location 2 exhibiting the highest completion rate. While discriminant function loadings (table 3.17 and figure 3.16) indicated the strongest discriminating ability for remoteness, average age and the percentage of Indigenous and Australian-born students, in terms of significant differences between their location means these categories were unremarkable. It is further worth considering that while institutes in location 1 displayed several traits that may be considered to have a negative connotation (such as the lowest completion rate, lowest percentage of graduates and lowest percentage of students enrolled in certificate III or higher courses), in respect of some outcomes these institutes scored exceedingly well. For instance, graduates of location 1 institutes had significantly higher satisfaction rates than students from other locations, and also attained significantly higher post-training salaries. Generally, the lack of a coherent association between the demographic, institutional and environmental variables on one side and combined institutional efficiency (for example, ‘teaching hours’ efficiency and ‘employment outcome’ efficiency)

indicates there are other factors, which we did not observe, that determine if an institute scores highly on both types of efficiencies. This means that, in the practical evaluation of the productivity in the vocational education sector it should thus be kept in mind that TAFE efficiency is a multidimensional concept and its results depend on carefully defined input and output measures. Efficiencies should be defined carefully depending on the specific property intended to be evaluated. In our study we defined two separate types of efficiency and created rankings for the TAFE institutes under examination. We found that efficiencies calculated under one definition are not necessarily an indicator for efficiencies obtained via alternative definitions. It therefore seems prudent to conclude that any results stemming from the efficiency analysis of Australian TAFE institutes, and by extension the efficiency of any group of public institutions, should always be accompanied by a carefully phrased explanation on how efficiency was specifically defined.

3.1.11 Conclusions

In this study we examined two types of TAFE efficiency, using two competing methods. The first efficiency type included the definition of a 'teaching hours' model, which evaluated the efficiency of the plain transformation of financial and administrative resources into teaching output. The second efficiency model brought a teaching quality-related aspect into the analysis. Here, we wanted to examine how efficiently TAFE institutes perform in achieving a measurable output, namely employment outcomes, for its students. In this context the second model could be viewed as not only examining pure efficiency, but also effectiveness. In addition to operationalising both models we wanted to know which demographic and institutional variables could be seen as predictors of the efficiencies we have estimated. We hypothesised that such knowledge might lead to actionable policy recommendations which could lead to increases in institutional efficiency. To accomplish this investigation we applied a technique that had not hitherto been used in the

evaluation of the Australian TAFE system: parametric Stochastic Frontier Analysis. We also aimed to apply a competing technique, non-parametric Data Envelopment Analysis, to the same models. This technique has been applied previously on the TAFE system, and our main interest was to evaluate to what extent both approaches yielded comparable results and to determine whether either of the two approaches was preferable to answer our research questions. The data necessary for this investigation came from a variety of sources. First, the prerequisite financial figures were obtained from 2011 institutional annual reports, institutional websites and regulatory state authorities. Second, most of the data relating to student demographics and institutional variables were derived from the 2011 Student Outcomes Survey. Finally, statistics to calculate institutional completion and load pass rates were retrieved from NCVER's Students and Courses data base.

There are a number of findings from this study that significantly extend the knowledge about the efficiency in the TAFE sector. In the 'teaching hours' SFA model we found that total expenditure and the number of course offered by institution provided valid inputs into the model. A very strong non-idiosyncratic error component in the model indicated that there was significant inefficiency in the TAFE system. Furthermore, we found the presence of a stochastic production frontier. Several demographic and institutional predictors of inefficiency were identified. We also established that there was a strong nonlinear relationship between institute size and efficiency in the band of smaller institutions. We determined the minimum size for an institute so that its size was not an impediment to efficiency.

A second analogous analysis of the 'teaching hours' model used the DEA approach. Here results of DEA analysis delivered strong correlations with the preceding SFA results in terms of the Constant-Returns-to-Scale sub-model but not with the Variable-Returns-to-Scale sub-model. Institute remoteness

emerged as a predictor of inefficiency that parametric and non-parametric approaches had in common.

The second type efficiency analysed in this study was 'Employment outcome' efficiency, measuring how TAFE institutes compared in the conversion of teaching resources to labour market results for its students. SFA of this model also found substantial inefficiencies across the TAFE sector. A number of predictors of inefficiency were found. Subsequent DEA analysis showed that there was less agreement between the two methods in terms of predictors for inefficiency. Comparing parametric with non-parametric efficiency results replicated the strong relationship between SFA scores and DEA CRS scores. Unlike in the 'teaching hours' model, in the 'employment outcome' there was a significant correlation between SFA and DEA VRS, albeit much weaker than between SFA and DEA CRS.

Finally, we evaluated whether there was a relationship between the two different types of efficiencies analysed in this study. We found that there was no correlation between 'teaching hours' efficiency' and 'employment outcome' efficiency when using the SFA and DEA VRS approach, but a mild negative correlation when using the DEA CRS method.

Given these results it is worth considering what recommendations could be informed by this research. Both theoretical and practical recommendations emerge. Theoretical recommendations relate to the methods of SFA and DEA themselves and deal with the issues of their applicability when employed in efficiency analysis in the TAFE sector. For researchers studying the efficiency of public institutions such as TAFE institutes it is helpful to know that with respect to considering the methods evaluated here we could conclude that SFA and DEA deliver comparable results and both can be applied in the context of TAFE evaluation. There are nuances in terms of when either method might be preferable. For instance, in analyses where only very few DMUs are involved it

may be preferable to apply DEA, as the parametric nature of SFA requires a certain minimum of data points to yield reasonable estimates. If a researcher does not have intimate knowledge of either the DEA or SFA technique it may also be advisable to employ DEA, as this approach is easier to implement due to the absence of the requirement to specify a functional form of the efficiency frontier and thus less conceptual knowledge is necessary to operationalise the model. The usage of SFA on the other hand may be preferable in situations where explicit statistical hypothesis testing is required. The SFA approach could have an advantage if DMU data contained several outliers, as DEA would not be well suited to deal with such data.

Practical recommendations emanating from this research aim to deal with the direct impact that the research results presented here could have on the TAFE sector. In our 'teaching hours' model we observed increasing returns to scale, meaning that teaching outputs could be increased disproportionately with the increase of financial and administrative resources. We also found significant economies of scale effects. Specifically, smaller institutes with less than 2.4 million annual teaching hours could increase their efficiency by either growing or amalgamating with other smaller institutes. In addition we found that remoteness and proportion of students with English as a second language were the most significant predictors of inefficiency. When interpreting these results one needs to bear in mind that it is mainly those remote institutes which are smaller. In realistic terms it may be difficult to grow or to amalgamate. In such cases there are of course policy issues that go beyond the pure efficiency consideration. Such issues might be the improvement of access to vocational education for rural residents or Indigenous Australians. The 'employment outcome' model presented in this paper evaluated how institutes utilised their resources and transformed them into measurable outcomes for its students in the post-training labour market. Here we noted that it was remote institutions that were associated with decreased efficiency. Small and

very small sized institutes on the other hand were associated with increased employment outcome efficiency. It may be possible that the infrastructure of these institutes, their location and a smaller number of students gives them an edge when converting their resources into employment outcomes.

Finally, we found that teaching hours and employment outcome efficiency were not linearly related. These two efficiencies have been specifically defined for the purpose of this study. Their main purpose was to evaluate two different efficiency dimensions: our first model aimed to determine the efficiency to convert financial and administrative inputs into teaching hours output, and our second model was designed to measure the efficiency in the transformation of teaching resources into a tangible post-study outcome, namely the success of an institute's graduates in obtaining post-training employment. Theoretically, it is possible to define an almost infinite number of other efficiencies. We see from this research that there is not necessarily a relationship between different types of efficiencies. For policymakers it is therefore necessary to take a multi-dimensional approach that takes into account the various aspects of different approaches to the concept of efficiency when making policy decisions. This means it is necessary for each efficiency model to be specified to evaluate one particular aspect of TAFE efficiency only, under careful selection of the appropriate variables that define the production function.

4 Completion in VET – How beneficial is it, and who completes? Portfolio paper 3

4.1 Structure

4.1.1 Introduction

In recent years there has been an increased policy focus on qualification completions in both VET and higher education sectors. Governments are keen to know the extent to which students complete the qualifications or studies they commence because they want to ensure adequate accountability for the expenditure of any public training funds that may have been used. They are also keen to understand how institutions measure up to national or state

training key performance objectives and targets. Educational institutions also want to better understand the extent to which their training programs meet these benchmarks for internal monitoring and quality assurance purposes. Individual student and employer consumers want to better understand how institutions are performing so that they can make informed decisions about the programs they should purchase.

In the university sector it has long been recognised that the completion of an undergraduate degree imparts significant social and economic benefits on students and society as a whole (Borland, 2001; James, 2001; Marks, 2007; Coates & Edward, 2009). As a consequence of the perceived and real benefits of obtaining a university degree, completion rates have been recorded at a relatively high level of 80%, along with an improving trend (ACER, 2011). Completions in vocational education and training (VET)[1] on the other hand have been known to be substantially lower than in the university sector. A variety of reasons have been cited to account for the low VET completions in Australia, including problems with employment and course, health and chance events, institutional factors, and financial and family problems (Long et al. 1995). Other authors concluded from the analysis of the English higher education system that wrong choice of field of study, financial difficulty and dissatisfaction with the training experience (Yorke et al., 1997), as well as inability to cope with the training demands, social dissatisfaction and dissatisfaction with the institutional environment (Yorke, 1999) contribute to significant non-completion.

One substantial dissimilarity between the completion of a qualification (or degree) in the university and VET sector context appears to be the perception of non-completion. In the university sector non-completion is often (but not

[1] Completion rate in the VET sector as used in the paper refers exclusively to course completion rate, unless otherwise specified. There are also module completion rates, which refer to the completion of individual modules. These completion rates do not exhibit the same definition problem as course completion rates, and are calculated simply by (modules completed divided by modules commenced)*100.

always) associated with failure, while in the VET sector it is generally accepted that non-completion frequently means that students have achieved the specific training goal which prompted them to enrol in a particular program (McInnes et al., 2000; Karmel & Nguyen, 2006; Mills et al., 2011). For instance, a plumber may enrol in a certain module to acquire the skills involving a new technique without the intention of gaining credit toward a qualification. Cohen and Brawer (1996) commented that to vocational students who are seeking a job, completing their training becomes irrelevant as soon as a job becomes available. There is thus less negative stigma attached to categories such as 'graduate' or 'dropout'. This, in addition to the above-mentioned impediments, may account for the considerable differences in completion rates and additionally necessitates caution when attempting to directly compare completion rates between both sectors, as these numbers need to be interpreted differently.

An additional issue in assessing the difference in completion rates between the university and VET sector is that there has been some controversy about the way in which completion rates are determined in the vocational system. At the current point in time there is no unique student identifier in the VET system and the concept of commencement of a qualification is not very well defined in the VET system (Mark & Karmel, 2010). Completion rates can therefore not be calculated in a straightforward manner. In their paper, Mark and Karmel therefore used a modelling technique in an attempt to derive completion rates in Australian vocational education. Their method used a Markov chain approach which has the property that the transition probabilities from one year to the next are not dependent on past transitions (this contrasts with more conventional time series methods such as ARIMA or vector autoregressive models where future predictions are, at least partially, a function of prior events). This technique enabled them to estimate completion rates for students commencing in 2005 and yielded an overall completion rate of 27.1%,

with significant variation between several categories of students (for example, full-time, part-time, age groups, course level). Bednarz (2012) performed an analysis using a similar methodology on students commencing in 2008 and estimated an overall completion rate of 28% for this commencing year. The results of these analyses have been met with some scepticism. Ross (2011) cited various industry figures who pointed to substantially higher completion rates in selected fields. For instance, an independent analysis by Service Skills Australia (SSA) yielded completion rates substantially higher than the published Mark and Karmel figure. Reasons cited as responsible for the discrepancies were temporal withdrawals such as for pregnancy, illness, relocation etc. which were not accounted for in the Mark and Karmel study. The authors of the original study dismissed these objections and suggested that the SSA study relied on cherry-picked training providers and on too short a period of analysis (Ross, 2011). Despite these somewhat conflicting research results with respect to actual completion figures, there can be little doubt that completion rates are a key indicator of TAFE performance. While there is some disagreement among commentators about the actual quantification of completion rates, there is a general perception that Australian VET completion rates are low (Mark & Karmel, 2010; Azemikhah, 2009; Snell & Hart, 2007).

In order to investigate whether students do commence their VET studies with the aim of not completing them, NCVER designed and conducted the Student Intentions Survey (SIS) in 2011 (NCVER, 2011c). This survey was intended to canvas completion intentions along with the collection of social and educational data. The main surprising result of this survey was that a very high percentage of students (93%) set out to complete their qualification. While this figure may have been somewhat inflated (it appears reasonable to speculate that some students may not have been willing to divulge their true intentions as, for instance, government funding may have depended on their declared aim to complete a prescribed course of study), it appeared

worthwhile to explore the unexpected large disparity between intended and actual completions.

It can be assumed that students make an individual choice when deciding to discontinue the path to a qualification. To make informed decisions, it is necessary that students consider the costs and benefits of completing or discontinuing their studies. It is therefore indicative to quantify the benefit of completion. Such evaluations have been done in the past (see Ryan, 2002; Karmel & Nguyen, 2006). These analyses have compared cohorts of completers and non-completers with respect to training outcomes for these groups (employment, salary etc.). We want to add to this body of research by analysing outcomes for individuals based on them either completing or not completing.

The purpose of this paper is therefore to evaluate a range of questions that surround contemporary issues about completion in vocational education. These questions can be broadly divided into two sub-themes. The first theme deals with the benefit that completion of a qualification imparts on the student. In this context we set out to investigate:

- What predictors are there for actual completion and how useful are they?
- What is the benefit of completing a qualification?
  - Does the benefit (and if so how) vary for different training outcomes (for example, employment, salary, further study etc.)
  - How is the benefit distributed across the student body?
  - Which student categories benefit most (and least)?

The second theme deals with the probability of completion itself. Here we asked:

- What predictors are there for intended completion and how useful are they?
- Is there a relationship between intended and actual completion?
- Can a typology be developed for those who intend to complete, but do not?

Ultimately the purpose of this research is not only the development of quantitative tools to answer the above research questions, but also to examine

current issues surrounding completion in the VET sector, and to give administrators in VET tools for addressing completion issues at their institutions. For instance, applying the tools presented in this paper may aid them in identifying potential non-completers. These students could then be approached and apprised of the benefit they would derive if they would complete their studies and thus be encouraged to re-evaluate their intentions.

4.1.2 The benefit of completion

There are two distinct types of completion rates in the VET sector, module completion rates and course completion rates. While the calculation of module completions is straightforward (they generally begin and finish within the same reporting period), the calculation of course completions — our main concern in this study — is much more complex. This is partially due to the lack of a unique student identifier, which makes it difficult to link a specific enrolment to a given completion across more than one reporting period.

Research in this area has been scant and only a few studies have attempted to develop a solution for this problem (Mark & Karmel, 2010; Bednarz, 2012). There has been controversy (Ross, 2011) as not all commentators agreed with the dismal overall estimated completion rates of 27% and 28% published in these studies. One possible issue with the work done by Mark and Karmel is the assumption that the probability of attrition of individual students remains constant over the period of their enrolment. We speculate that this assumption is not necessarily correct. The problem of the lack of universally accepted completion rates in the VET sector will persist until the introduction of a universal student identifier removes the main obstacle to transparently calculating such figures[2].

[2] The well-known problem of a unique student identifier is currently being addressed and its introduction is being planned for 2015 (DET, 2014).

To avoid entering the methodological debate with respect to the appropriate method of their calculation, in this study we used a completion definition that arose out of the collection of data for the Student Outcome Survey (SOS). The SOS is an annual survey (with alternating 'small' and 'big' sample sizes in order to satisfy different reporting levels in odd and even years[3]) that focuses on student outcomes and satisfaction with VET and also collects data on personal and training characteristics, employment outcomes, further training activity, satisfaction with the training, whether the main reason for undertaking the training was achieved, whether the training was relevant for the current job and reasons that led to the discontinuation of the training (if applicable). The survey categorises students into two discrete groups, Graduates and Module completers, defined as follows: Graduates are students who have completed a qualification through their training, including a bachelor's degree or higher, an advanced diploma or diploma, or any certificate from I to IV. Module completers on the other hand are students who have successfully completed part of a course (for example, at least one module) without completing a qualification and who have left the vocational education and training system by the time the survey was undertaken (NCVER, 2011b). We used this definition to identify completers (Graduates) and non-completers (Module completers) in our analysis. The benefit of utilising this categorisation was that a large amount of data was readily available for analysis and that the categorisation into these two groups was largely uncontroversial and accepted in the VET community. The primary drawback of this approach was that it omitted those students who enrolled but have not completed a single module. However, with module completion rates of around 90%, we assumed that this group of students would have accounted for only a very small percentage of non-completing students. Furthermore, the total

[3] For a more in-depth description of this survey please see paper 1 of this portfolio.

absence of any data for this type of student at the national level would have made their inclusion in any analysis problematic.

The benefit that we aimed to investigate in this study is the quantification of the ratio of the probability (or predicted value) of the outcome of a number of post-training indicators based on completion and non-completion. Benefit is thus conceptualised as (1):

$$Benefit = \frac{Prob[outcome_{completer}]}{Prob[outcome_{non-completer}]} \quad (1)$$

In the case of outcomes that are estimated via OLS regression (for example, salaries, satisfaction) the benefit is calculated as the ratio of the predicted outcome under a completion and non-completion scenario (2).

$$Benefit = \frac{Pred[Outcome_{\ completer}]}{Pred[Outcome_{non-completer}]} \quad (2)$$

If, for instance, an individual has a probability of being employed after training as a completer of 0.8 and a probability of being employed after training as a non-completer of 0.8 then this individual would be deemed to have drawn no quantifiable benefit of completion in terms of employment as the benefit equals one. A different individual with probabilities of 0.9 and 0.8 respectively would be considered to draw a benefit of 1.125, or 12.5%, provided he/she completed their qualification. The indicators we evaluated were post-training employment, salary, occupational status, satisfaction with the training, further study after training and improved employment conditions after training. These indicators were derived directly from student responses to the survey except for the occupational status indicator which was created on the basis of the Australian Socioeconomic Index, as suggested by McMillan, Beavis, and Johnes (2009).

4.1.2.1 Data and methods

We opted for the 2011 wave of the Student Outcome Survey as the year for which we conducted our analysis. The reasons for this were mostly dictated by

the administrative schedule in which the survey is conducted: 2011 represented the most recent year for which a complete data set was available that surveyed students with the intention of institutional reporting (for example, a 'big' year). This survey was conducted on a sample of around 300,000 students, of which around 110,000 responded. Of these 110,000 respondents, around 46,000 were graduates, 31,000 were module completers, and 33,000 were continuing students who were surveyed under the assumption they had left the VET system. However, their response to the survey indicated their continued enrolment. We thus discarded continuing students from our analysis. A further reason for the choice of 2011 as the base year for analysis was that the Student Intention Survey was conducted in that year and we intended to link these two surveys for later analysis.

To evaluate the benefit of completion we took the following approach. First we described the two distinct groups of completers and non-completers with respect to demographic, educational and institutional variables. Then, we developed a model which predicted a particular kind of outcome, for instance, employment after training. This model was first run over the subset of completers only; however, we predicted the probabilities of employment for completers and also for non-completers. Next, we ran a model with identical predictors over the subset of non-completers, and predicted employment probability for non-completers and completers. As a result, our dataset then contained two probabilities for each individual student, the first for the completer scenario and the second for the non-completer scenario. The ratio of the two probabilities gave us an estimate of the individual benefit. A histogram of the computed benefit value was created and displayed how the benefit was distributed across the student body. A final step in evaluating the benefit of completion was the creation of a decision tree diagram via the chi squared automated interaction detection (CHAID) method. This method was developed by Kass (1980) and partitioned the data into mutually exclusive,

exhaustive subsets that best predicted the outcome. The method proceeded in several steps. In the first step, the best partition for each predictor is found. Then the predictors were compared and the strongest predictor was chosen. The data were then subdivided according to this predictor. All subgroups were then re-analysed independently to produce further subdivisions (Kass, 1980). The advantage of this method was that data could be readily and hierarchically partitioned into successively more homogenous groups, the patterns for which could be displayed via an easily interpretable decision tree.

We first took a look at the variables that were directly related to the specific outcomes we set out to investigate in this study (table 4.1).

Table 4.1 Outcome related variables

		Graduates	Module completers	Total	p{Diff}
Study status	Part-time	27.8%	99.0%	68.1%	<0.01
	Full-time	72.2%	1.0%	31.9%	<0.01
Employed after training	Not employed	21.3%	24.5%	23.1%	<0.01
	Employed	78.7%	75.5%	76.9%	<0.01
Further study after training	No further study	65.4%	86.0%	77.0%	<0.01
	Further study	34.6%	14.0%	23.0%	<0.01
Annual earnings (full time)	mean	51,874	59,161	55,800	<0.01
Improved employment	Improved	63.6%	49.1%	55.5%	<0.01
Condition after training	Not improved	36.4%	50.9%	44.5%	<0.01
Satisfaction with training	mean	4.27	4.15	4.20	<0.01
Occupational status	mean	43.23	44.03	43.65	<0.01

P{Diff} indicates significance of difference between Graduates and Module completers.
Test employed were Chi-squared for categorical and t-test for nominal variables

As one would have expected, student enrolment status differed substantially between completers and non-completers. Almost three quarters of completers had enrolled full time, compared with wide-ranging part-time enrolment of non-completers. Completers also engaged in further studies after their training in far greater proportions than non-completers. This also appeared intuitive, as it

can be assumed that completion of the program was a necessary prerequisite for several additional training pathways. Surprisingly, the post-training salaries appeared to be significantly higher for non-completers than completers. However, these overall salary figures may be misleading, as non-completion may have been particularly prevalent in certain high-income fields such as mining. Completers seemed to be more satisfied with their training but their post-training occupational status fell somewhat short of non-completers. However, completers gained more occupational status as a result of their training.

In respect to student demographics (table 4.2) males appeared more likely to be non-completers whereas females were slightly more likely to be completers. Younger age groups tended to be more likely to be completers as compared with older age groups.

There were only minor differences in completion across socioeconomic groups. Non-completion seemed to be more prevalent in remote and regional areas than in urban areas.

Table 4.2 Demographic, educational and institutional variables

		Graduates	Module completers	Total	p{Diff}
Sex	Males	45.9%	55.9%	51.5%	<0.01
	Females	53.9%	43.8%	48.2%	<0.01
Age group	15–24	42.8%	27.1%	33.9%	<0.01
	25–44	37.7%	39.3%	38.6%	<0.01
	45+	19.5%	33.7%	27.5%	<0.01
SEIFA	Most disadvantaged	14.0%	14.8%	14.4%	<0.01
	Very disadvantaged	23.8%	25.8%	24.9%	<0.01
	Somewhat disadvantaged	22.3%	22.4%	22.4%	
	Little disadvantaged	22.7%	21.2%	21.9%	<0.01
	Least disadvantaged	17.2%	15.8%	16.4%	<0.01
Remoteness	City	59.5%	50.1%	54.2%	<0.01
	Regional	37.3%	44.1%	41.1%	<0.01
	Remote	3.2%	5.8%	4.6%	<0.01
Prior education	Diploma & above	21.6%	27.4%	24.8%	<0.01
	Cert III & IV	24.4%	23.2%	23.7%	<0.01
	Y12/below or Cert I/II	54.0%	49.4%	51.5%	<0.01
Qualification level	Cert I & II	18.6%	22.6%	20.9%	<0.01
	Cert III & IV	62.7%	32.8%	45.8%	<0.01
	Diploma & above	18.7%	5.4%	11.2%	<0.01
	Other	0.0%	39.2%	22.2%	<0.01
Employed before training	Employed	72.4%	73.1%	72.8%	<0.01
	Unemployed	15.1%	12.9%	13.9%	<0.01
	Not in labour force	12.5%	14.0%	13.4%	<0.01
Field of education	Natural Sciences	0.7%	0.3%	0.5%	<0.01
	Information Technology	1.9%	2.9%	2.5%	<0.01
	Engineering	16.0%	15.5%	15.7%	<0.01
	Architecture	6.2%	6.5%	6.4%	<0.01
	Agriculture	3.1%	6.6%	5.1%	<0.01
	Health	6.0%	6.7%	6.4%	<0.01
	Education	6.3%	3.6%	4.8%	<0.01
	Management	29.4%	14.0%	20.7%	<0.01
	Society & Culture	16.7%	5.9%	10.6%	<0.01
	Creative Arts	3.1%	2.4%	2.7%	<0.01
	Food and Hospitality	5.9%	11.0%	8.8%	<0.01
	Mixed Field	4.7%	14.9%	10.4%	<0.01
	Subject Enrolment	0.0%	9.6%	5.4%	<0.01

P{Diff} indicates significance of difference between graduates and module completers.
Test employed were chi-squared for categorical and t-test for nominal variables

The benefit of completion with respect to employment

The first of our outcome indicators for which we analysed the benefit of completion was employment after training. The model we developed to predict

employment after training consisted of the following independent variables: sex (male/female), age group (15—24, 25—44, 45 and above), socioeconomic status (most disadvantaged, very disadvantaged, somewhat disadvantaged, little disadvantaged and least disadvantaged; as per ABS Socioeconomic Index For Area (SEIFA, ABS, 2013), field of education (technical, business, community services, other services, other), remoteness (city, regional, remote; as per ABS Standard geographical classification (ABS, 2013b), study mode (full-time, part-time), employment status before training (employed, unemployed, not in labour force), prior education (diploma and above, certificate I and II, certificate III and IV), and qualification level (diploma and above, certificate I and II, certificate III and IV). All categorical variables were dummy coded. The model statistics for the completer and non-completer model were fairly similar (table 4.3), although the non-completer model had a slightly higher explanatory power than the completer model (Nagelkerke Rsq 0.39 v 0.31). Coefficients were also comparable for completer and non-completer models. The strongest predictive power for obtaining post-training employment came from being in employment prior to the training. The influence was stronger for non-completers and may have resulted from the high proportion of part-time students among this group. It is most likely that many students of this group were holding on to their employment while studying.

Table 4.3 Employment after training model

Employed after training		Graduates			Module completers		
		B	Sig.	Exp(B)	B	Sig.	Exp(B)
Sex	Female	-0.08	<.01	0.93	-0.18	<.01	0.84
	Male	0.00					
SEIFA	Most disadvantaged	-0.50	<.01	0.61	-0.27	<.01	0.76
	Very disadvantaged	-0.29	<.01	0.75	-0.25	<.01	0.78
	Somewhat disadvantaged	-0.21	<.01	0.81	-0.09	<.01	0.91
	Little disadvantaged	-0.10	<.01	0.90	0.00	0.91	1.00
	Least disadvantaged	0.00			0.00		
Field of Education	Technical	0.72	<.01	2.06	0.32	<.01	1.37
	Business	0.30	<.01	1.34	0.06	<.01	1.06
	Comm. services	0.22	<.01	1.25	0.00	0.89	1.00
	Other services	0.63	<.01	1.88	0.42	<.01	1.52
	Other	0.00			0.00		
Remoteness	City	-0.50	<.01	0.61	-0.38	<.01	0.69
	Regional	-0.20	<.01	0.82	-0.14	<.01	0.87
	Remote	0.00			0.00		
Age group	Age 15–24	0.21	<.01	1.24	0.15	<.01	1.16
	Age 25–44	0.10	<.01	1.11	0.32	<.01	1.37
	Age 45+	0.00			0.00		
Study mode	Part-time	0.18	<.01	1.20	0.38	<.01	1.47
	Full-time	0.00			0.00		
Status before training	Employed	2.30	<.01	9.97	2.78	<.01	16.09
	Unemployed	0.24	<.01	1.27	0.59	<.01	1.81
	Not in labour force	0.00			0.00		
Prior education	Diploma and above	0.03	0.04	1.03	0.35	<.01	1.42
	Cert III/IV	0.04	<0.01	1.05	0.35	<.01	1.41
	Cert I/II	0.00			0.00		
Qualification level	Diploma and above	0.52	<.01	1.68	0.34	<.01	1.41
	Cert III/IV	0.63	<.01	1.88	0.29	<.01	1.34
	Cert I/II	0.00			0.00		
	Constant	-0.54	<.01	0.58	-1.31	0	0.27
		Pseudo rsq		0.31	Pseudo rsq		0.39
		Chi sq = 72829.9 (p<0.01)			Chi sq = 77313.6 (p<0.01)		

Notes: In this and the following tables, B indicates regression coefficients, Sig indicates significance level, and Exp(B) displays e (2.72) raised to the value of the regression coefficient and represents the odds ratio in relation to the base category. As an example, in this table Exp(B) of 0.93 for female graduates signifies a 7% lower likelihood of females to be employed after training compared to males.

The running of both models and predicting the probability of post-training employment for each individual based on the completion and non-completion scenario enabled us to calculate the individual completion benefit for every student and create a histogram (figure 4.1).

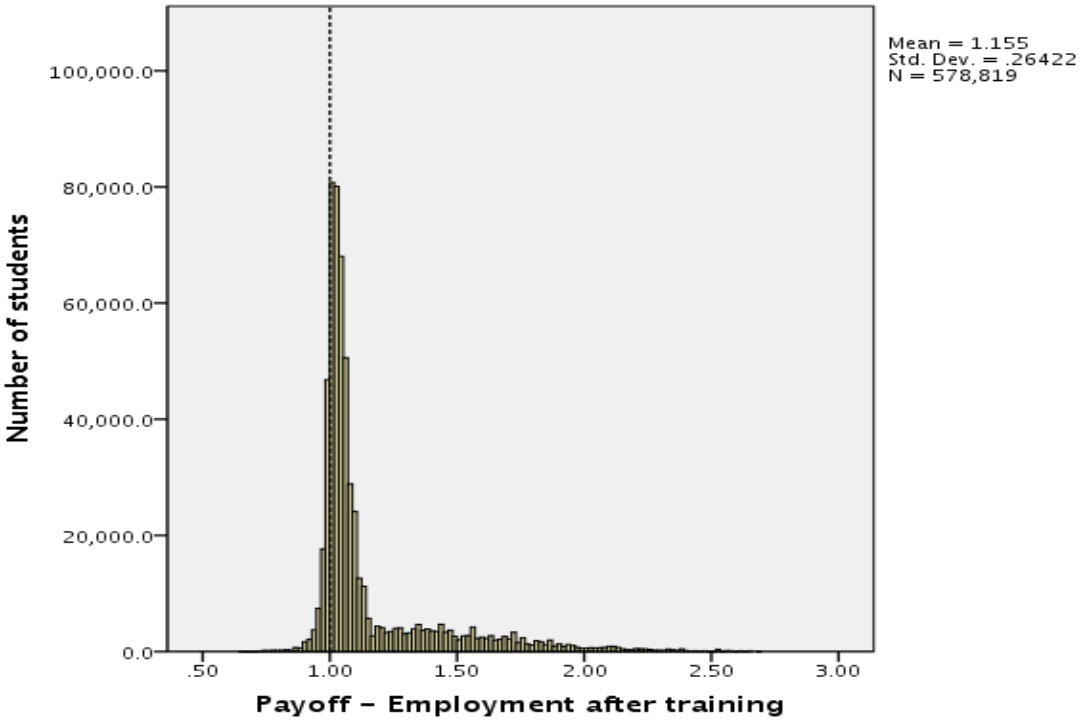

Figure 4.1 Histogram of completion benefit for 'Employment after training'

The mean completion benefit was 1.155, indicating that completing a qualification conferred a 15.5% advantage over a non-completer in terms of the probability of gaining post-training employment. It is thus beneficial to consider completion when enrolled in a VET course as completion does have a measurable benefit in terms of employment after training. In figure 4.1 we added a vertical line at the value of one. We noted that there was a small group of students (14.2%) to whom completion did not impart a benefit (for example, those to the left of 1.0) (table 4.4). Then, immediately to the right of 1.0, there were a large number of students for whom completion was only marginally beneficial. However, overall 85.8% of students received a positive return from completing their qualification.

Table 4.4 Payoff summary 'Employment after training'

	Employment after training	
	Number	Per cent
Payoff > 1	496,490	85.8
Payoff < = 1	82,328	14.2

It could then be asked: Who exactly were the students for whom completion mattered so much? We used the completion benefit scores to perform a Chi-squared Automated Interaction Detection, using an exhaustive CHAID model with the predictors of our regression model. This technique would not only give us a ranking of the most important predictors of completion benefits, but also help us to identify those subgroups of students for whom completion matters most (and least).

The resulting graph can be seen in figure 4.2. At the top level we noted the overall completion benefit of 1.155. The next level in the tree indicated the most significant predictor of the benefit of completion for the gaining of employment after training. The mean value for the individual categories signified the benefit for this group. In this outcome category, those who were not in the labour force before training derived the strongest benefit (1.67) from completion. These students were thus 67% more likely to be in post-training employment if they completed than if they did not complete. Staying in the 'not in labour force before training' category and moving another level down it could be seen that full-time students derived a benefit of 1.858 from completion whereas part-time students derived a benefit of 1.548. At the lowest level of the tree it could be seen that those with a prior education of Year 12 or below or Certificate I or II derived a benefit of 2.018 from completion. It was thus clear that a student who completed his or her qualification and who was not in the labour force before training and studied full-time and had relatively low prior qualification had more than twice the probability of obtaining post-training employment. On the other hand, someone who was in employment prior to the training and studied part-time derived only a relatively modest benefit of 1.8% from completion.

In the CHAID tree we visualised only those categories that differed substantially from others (for instance, we collapsed all categories where the

split yielded only small differences[4]), but it is clear that the completion benefit can be quantified for every possible student category. This underlined not only the theoretical but also the practical usefulness of this method in identifying those student categories for which completion matters most and least.

[4] In this and the following CHAID diagrams, a small '+' in the bottom right corner of a branch endpoint indicates that subsequent splits differed only marginally and have thus been collapsed.

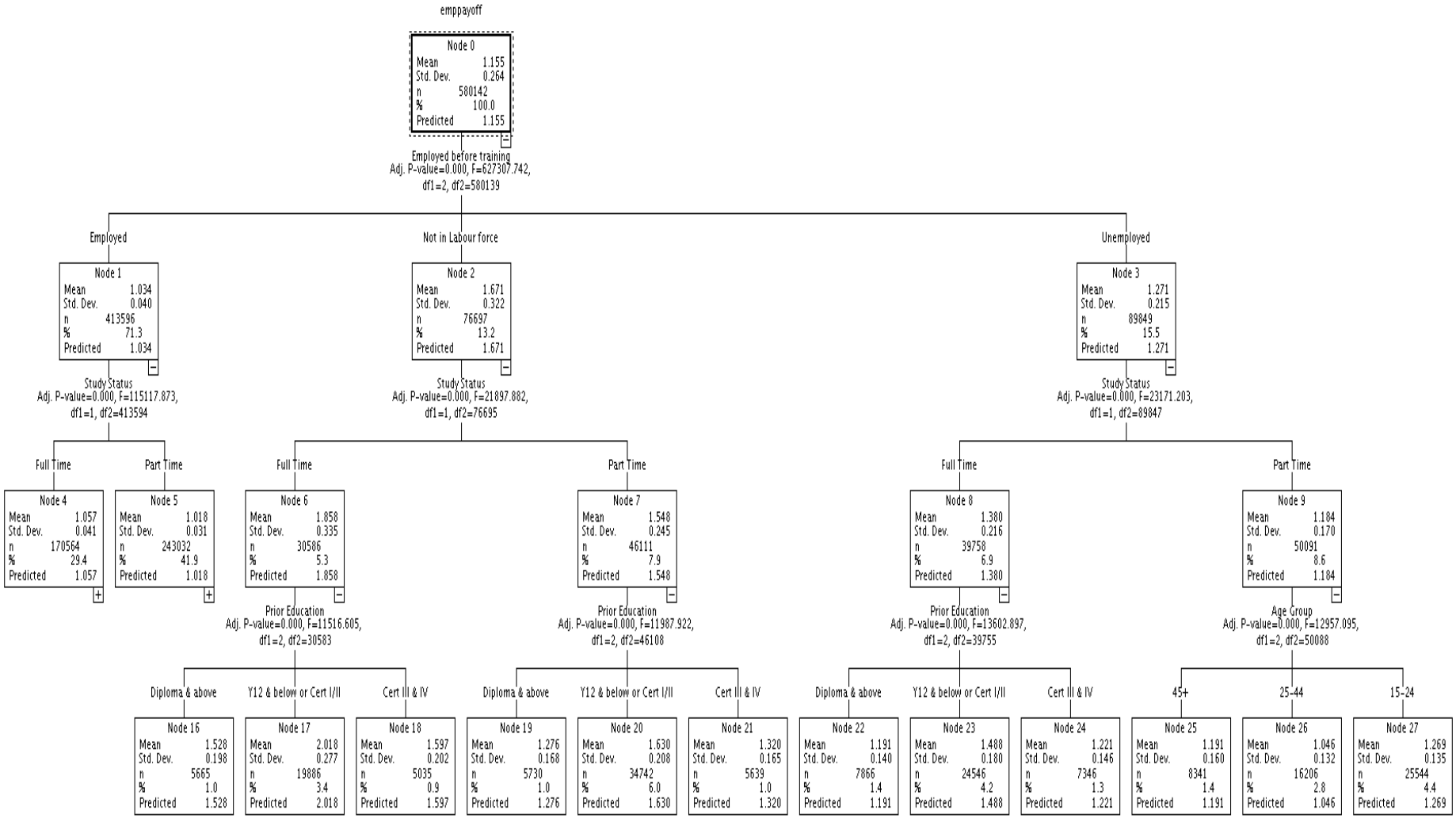

Figure 4.2 CHAID diagram – Completion benefit of 'Employment after training'

The benefit of completion with respect to salaries

In addition to the impact of completion on post-training employment prospects, VET students may be particularly interested in the impact of completion on their post-training salaries. We conducted our analysis using a similar methodology with this outcome variable. The main difference from the employment outcome analysis was that in terms of post-training salaries we deal with a quasi-continuous variable[5] instead of a dichotomous outcome. We thus employed ordinary least squares regression to predict salaries for completers and non-completers. A further difference to the employment model is the substantially lower response rate that the SOS receives for the salary question. While those who respond to the SOS usually complete the large majority of individual questions, the salary question receives only about a third of the response rate of other SOS items. Results of the salary models for completers and non-completers can be seen in table 4.5.

[5] The Student Outcome Survey collects salary data as the responses to a number of proposed salary ranges. This has been found to increase the number of responses to this survey question. We determined the mid-point of these ranges and used them in our regression analysis.

Table 4.5 Full-time salary after training regression models

Full time salary after training		Completers			Non-completers		
		B	Sig.	Beta	B	Sig.	Beta
Sex	Female	-8,412.3	<.01	-0.18	-7,713.1	<0.01	-0.14
	Male	0.0			0.0		
SEIFA	Most disadvantaged	-3,837.7	<.01	-0.05	-6,518.4	<0.01	-0.09
	Very disadvantaged	-3,688.3	<.01	-0.07	-5,214.9	<0.01	-0.09
	Somewhat disadvantaged	-1,695.2	<.01	-0.03	-3,457.6	<0.01	-0.06
	Little disadvantaged	-1,395.7	<.01	-0.03	-663.1	<0.01	-0.01
	Least disadvantaged	0.0			0.0		
Field of Education	Technical	8,092.6	<.01	0.17	8,966.4	<0.01	0.18
	Business	5,170.3	<.01	0.10	4,351.2	<0.01	0.07
	Comm. services	-1,450.8	<.01	-0.02	1,396.4	<0.01	0.02
	Other services	7,832.7	<.01	0.12	5,208.0	<0.01	0.09
	Other	0.0			0.0		
Remoteness	City	-12,290.5	<.01	-0.26	-13,100.6	<0.01	-0.27
	Regional	-10,870.0	<.01	-0.23	-11,274.0	<0.01	-0.23
	Remote	0.0			0.0		
Age group	Age 15–24	-16,857.0	<.01	-0.36	-16,867.3	<0.01	-0.29
	Age 25–44	-3,654.8	<.01	-0.08	-2,904.1	<0.01	-0.06
	Age 45+	0.0			0.0		
Study mode	Part-time	-616.5	<.01	-0.01	-530.0	0.36	0.00
	Full-time	0.0			0.0		
Status before training	Employed	3,973.5	<.01	0.06	6,517.5	<0.01	0.08
	Unemployed	-3,900.2	<.01	-0.04	-637.9	0.13	-0.01
	Not in labour force	0.0			0.0		
Prior education	Diploma and above	9,606.2	<.01	0.17	13,351.6	<0.01	0.25
	Cert III/IV	4,130.6	<.01	0.08	7,638.2	<0.01	0.14
	Cert I/II	0.0			0.0		
Qualification level	Diploma and above	13,154.8	<.01	0.23	3,577.4	<0.01	0.04
	Cert III/IV	8,401.9	<.01	0.17	1,459.5	<0.01	0.03
	Cert I/II	0.0			0.0		
	Constant	56,664.5	<.01		59,896.5	<0.01	
		Adj. rsq		0.30	Adj rsq		0.26
		F = 3263.6 (p<0.01)			F = 1852.9 (p<0.01)		

The pattern of post-training salary predictors is roughly the same between completers and non-completers. Higher pre-training qualification, pre-training employment, higher SES, male sex, technical and other services field of education, age >25, remote location and the qualification level of the training were all related to higher salaries for completers as well as non-completers. Both models had a reasonable explanatory value, in the case of completers the

explained variance amounted to 30% and for non-completers slightly lower at 26%. Calculation of the benefit yielded an overall mean of 0.983. The histogram is displayed in figure 4.3.

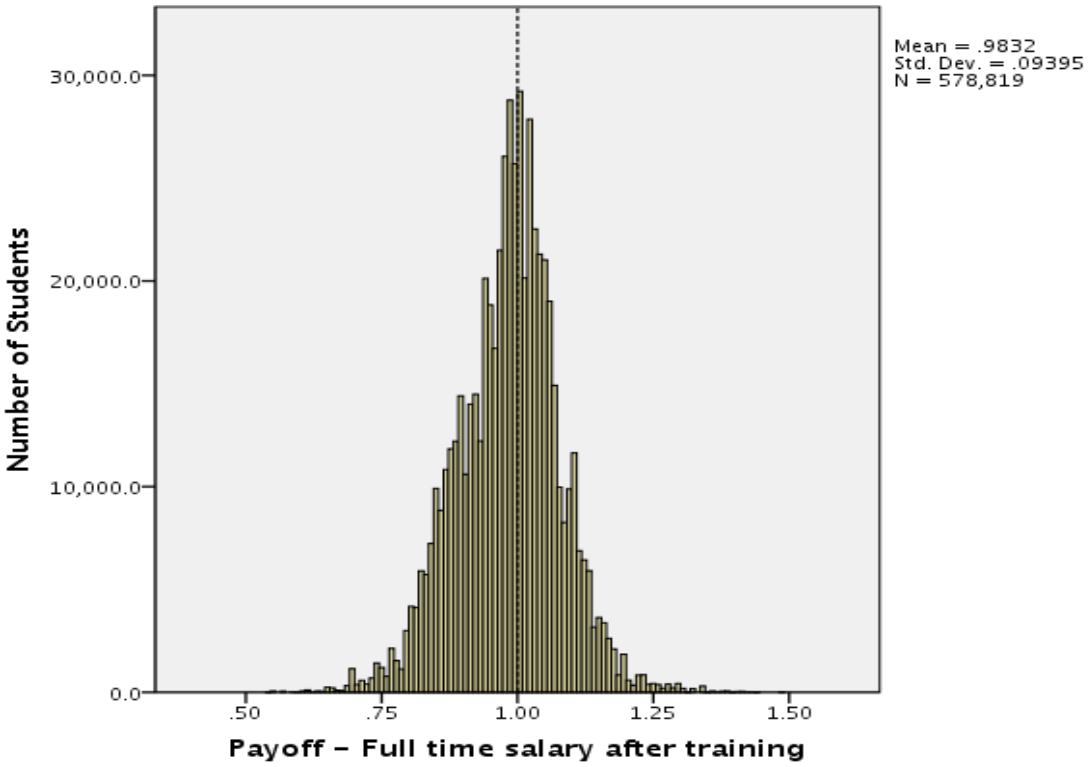

Figure 4.3 Histogram of completion benefit for full time salaries

The mean of the completion benefit of post-training salaries was close to 1 (0.983), suggesting that there was no overall benefit to completion. The salary completion benefit histogram looked also substantively different from the histogram of the completion benefit to employment after training histogram (which was substantially positively skewed, recall figure 4.1). The histogram suggested that slightly fewer students (44.7%, table 4.6) benefited from completion than students who drew a negative benefit from completing their qualification (55.3%).

Table 4.6 Payoff summary 'Full time salary after training'

	Full time salary after training	
	Number	Per cent
Payoff > 1	258,708	44.7
Payoff <= 1	320,111	55.3

It appeared to be a surprising result that people should be disadvantaged in salary terms by completing their qualification. We created the CHAID diagram for the completion benefit to post-training full-time salaries (figure 4.4). We noted from the CHAID diagram that all student groups with a negative benefit to completion were nested under the Certificate I and II category while students in the Certificate III and IV and Diploma or higher had overall positive salary benefits to completion. While the negative benefit to completion for certificates I and II appeared to be counter-intuitive, it had been recognised before. Dismal pay-offs to completion of lower qualifications were also found by Karmel and Nguyen (2006), Ryan (2011) and Karmel (2011). Specifically Ryan (2011) found that the benefit of completing a Year 12 qualification outweighed the benefit of completing a lower-level VET qualification. On the other hand, Herault, Zakirova and Buddelmeyer (2011) observed a small benefit from completion for lower-level qualifications, although this benefit disappeared after the first year in the workforce. However, they did caution that their results may have been constrained by the small size of the sample they analysed.

The lack of benefit from completing a Certificate I or II qualification in terms of earnings underscored a general shortcoming with respect to the direct value of such a low-level qualification. Long and Shah (2008) analysed returns for VET students and found that the best returns materialised for those students who enrolled in Certificate III and higher qualifications. The lack of significant benefits may encourage students of lower-level qualifications to forego entering the labour market and utilise their qualifications instead to pursue further educational opportunities, including apprentice and traineeships from

which they may derive more tangible benefits. Oliver (2012) also observed that while lower-level VET qualifications lacked a direct labour market benefit, they served as a stepping stone to further study. He also found that it were predominantly students from more disadvantaged backgrounds who benefited in this way from such lower-level VET qualifications. Stanwick (2006) studied the outcomes for students enrolled in low-level qualifications and concluded that they had mixed labour market outcomes. Younger students (<19 year-olds) had better outcomes than older ones. Stanwick also reported the comparatively higher propensity of certificate I/II students to enrol in further study and that this effect was where the true benefit to completion resided for students in this category.

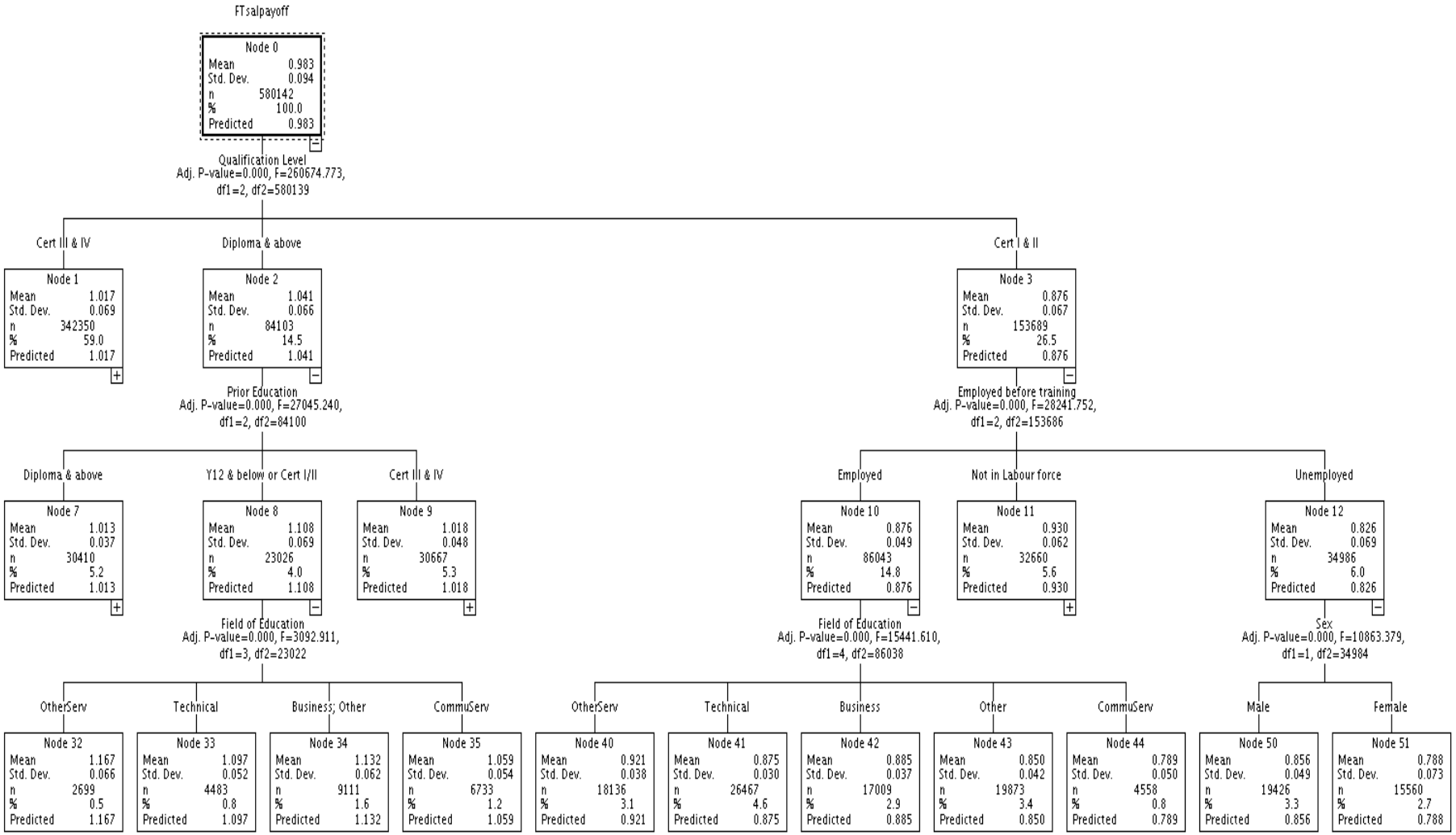

Figure 4.4 CHAID diagram – Completion benefit to 'Full time salary after training'

It was of interest to repeat the CHAID analysis specifically for the full-time salary payoff by investigating the salary outcome itself, that is, the difference between the predicted salary under the completing and non-completing scenario. While the result for the full-time salary difference could be expected to be fairly similar to the full-time salary payoff, the use of real life '$' figures could be seen as providing a more intuitive picture. The histogram for the completion salary difference can be seen in figure 4.5.

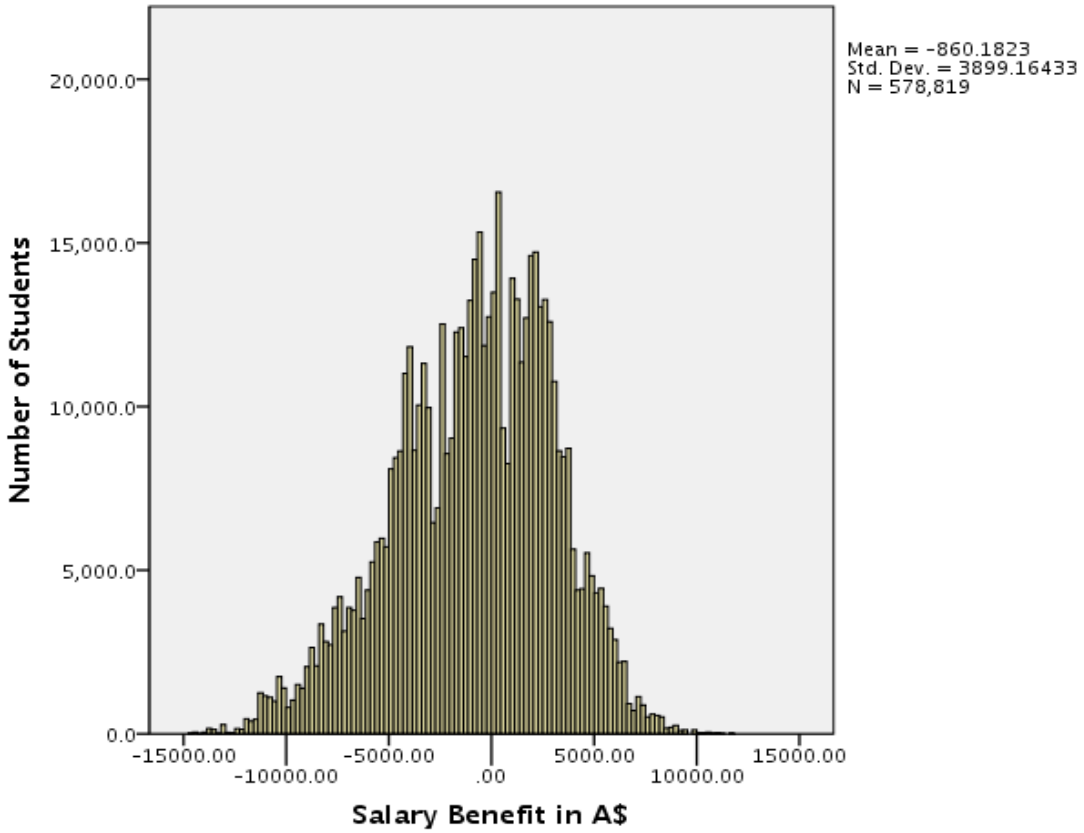

Figure 4.5 Histogram of predicted 'Full-time salary difference between completion and non-completion'

As expected, the distribution of the salary difference between completers and non-completers looks quite similar to the payoff histogram, along with the mean of the difference which was slightly negative at -$860. It is notable that the salary outcome to completion can potentially be quite negative and some values were estimated to be below -$10,000. The corresponding CHAID diagram (figure 4.6) resembled largely the CHAID diagram of the salary

payoff to completion (figure 4.4), but also gave an indication of the financial magnitude of difference in salary outcome from completion versus non-completion. As noted earlier, it was mostly lower qualifications (Certificate I + II) that were affected by a negative salary difference and outcomes were particularly abysmal for those who completed a Certificate I/II while having a prior education at a higher level (for example, Diploma: -A$8,318 and Certificate III/IV: -A$7,983).

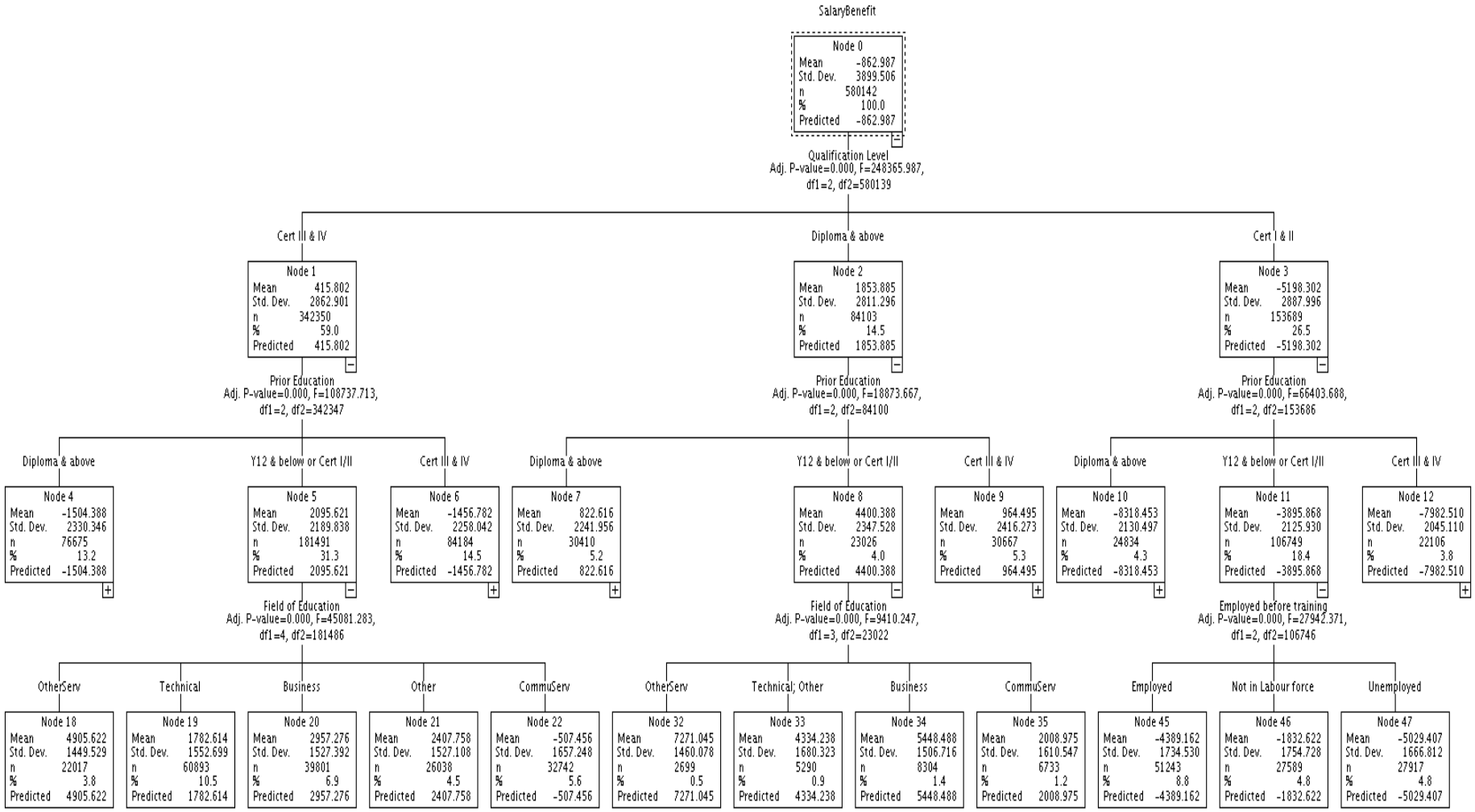

Figure 4.6 CHAID diagram – predicted 'Full-time salary difference between completion and non-completion'

The benefit of completion with respect to occupational status

Occupational status is one possible indicator of socioeconomic status (Hauser & Warren 1997). It is therefore useful to evaluate the benefit of completion in terms of occupational status. We assigned every student in post-training employment a value for occupational status based on his or her stated occupation classification. This value was derived using the methodology outlined in McMillan et al. (2009) and was located on a scale from 0 to 100 where increasing values indicated higher occupational status. Subsequently we employed two ordinary least squares regression models to predict occupational status based on completion and non-completion scenarios. The model estimates can be found in table 4.7.

Table 4.7 Occupational status after training regression models

Occupational status after training		Completers			Non-completers		
		B	Sig.	Beta	B	Sig.	Beta
Sex	Female	0.32	<0.01	0.01	2.31	<0.01	0.05
	Male	0.00			0.00		
SEIFA	Most disadvantaged	-2.28	<0.01	-0.04	-4.26	<0.01	-0.07
	Very disadvantaged	-2.00	<0.01	-0.05	-3.38	<0.01	-0.07
	Somewhat disadvantaged	-1.21	<0.01	-0.03	-2.94	<0.01	-0.06
	Little disadvantaged	-0.98	<0.01	-0.02	-1.03	<0.01	-0.02
	Least disadvantaged	0.00			0.00		
Field of Education	Technical	3.40	<0.01	0.08	1.50	<0.01	0.03
	Business	8.07	<0.01	0.19	5.88	<0.01	0.11
	Comm. services	4.16	<0.01	0.09	3.31	<0.01	0.05
	Other services	10.35	<0.01	0.21	4.40	<0.01	0.09
	Other	0.00			0.00		
Remoteness	City	-0.80	<0.01	-0.02	-0.66	<0.01	-0.02
	Regional	-1.10	<0.01	-0.03	-0.59	<0.01	-0.01
	Remote	0.00			0.00		
Age group	Age 15–24	-6.61	<0.01	-0.17	-5.92	<0.01	-0.13
	Age 25–44	-2.66	<0.01	-0.07	-2.40	<0.01	-0.06
	Age 45+	0.00			0.00		
Study mode	Part-time	0.06	0.406	0.00	1.35	<0.01	0.01
	Full-time	0.00			0.00		
Status before training	Employed	1.44	<0.01	0.03	2.80	<0.01	0.05
	Unemployed	-1.72	<0.01	-0.03	-1.05	<0.01	-0.01
	Not in labour force	0.00			0.00		
Prior education	Diploma and above	13.45	<0.01	0.30	20.68	<0.01	0.44
	Cert III/IV	3.14	<0.01	0.07	4.59	<0.01	0.09
	Cert I/II	0.00			0.00		
Qualification level	Diploma and above	12.00	<0.01	0.26	5.13	<0.01	0.07
	Cert III/IV	5.07	<0.01	0.13	-0.16	0.105	0.00
	Cert I/II	0.00			0.00		
	Constant	32.62	<0.01		33.19	<0.01	
		Adj. rsq		0.28	Adj rsq		0.28
		F = 4831.7 (p<0.01)			F = 3250.3 (p<0.01)		

Variables in our completer and non-completer models both explained a similar amount of variance (28%) in post-training occupational status. The strongest predictor in both models was the attained education level prior to the recent training (beta = 0.44 for non-completers and beta = 0.3 for completers). Female sex, older age groups, remote location, enrolment at the diploma and above level, and enrolment in business, technical, community and other services programs were also associated with higher occupational status of completers and non-completers. The predicted occupational status scores for

both categories could then be used to derive the individual benefit from completion. The resulting histogram is provided in figure 4.7.

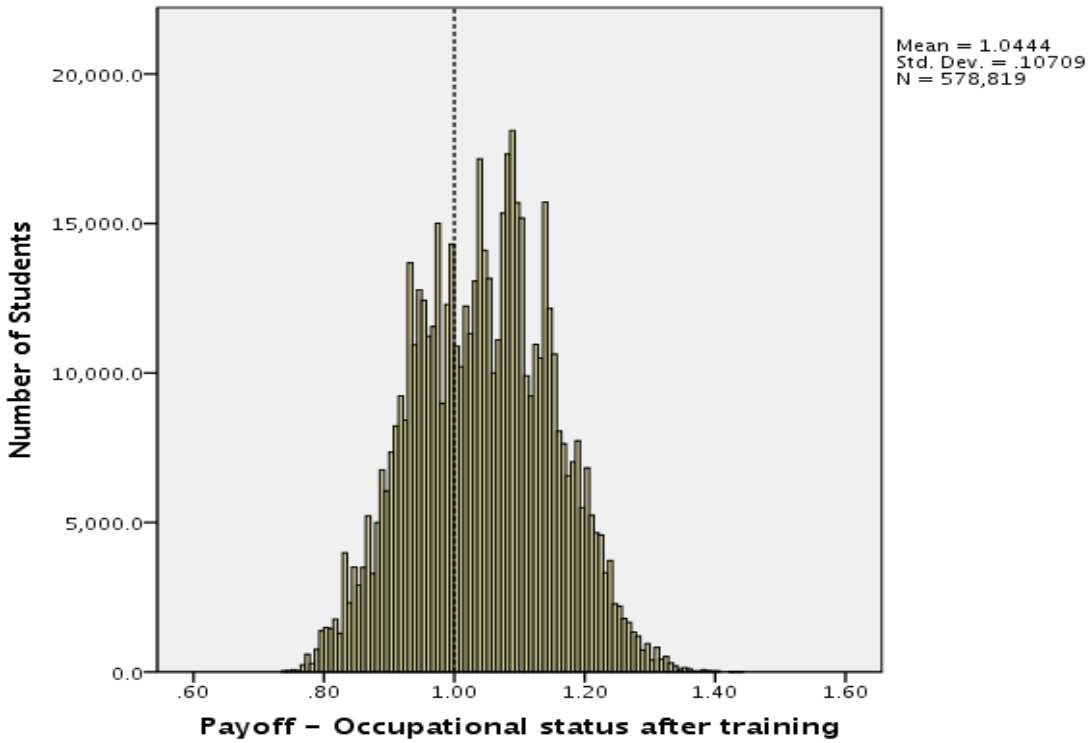

Figure 4.7 Histogram of completion benefit for occupational status

The histogram indicates that there is a modest overall benefit to completion of 4.4%. A significant majority of students (64%) clearly benefited from completion although about a third of all students did not derive any benefits from completion in terms of occupational status (table 4.8).

Table 4.8 Payoff summary 'Occupational status after training'

	Occupational status after training	
	Number	Per cent
Payoff > 1	370,363	64.0
Payoff <= 1	208,456	36.0

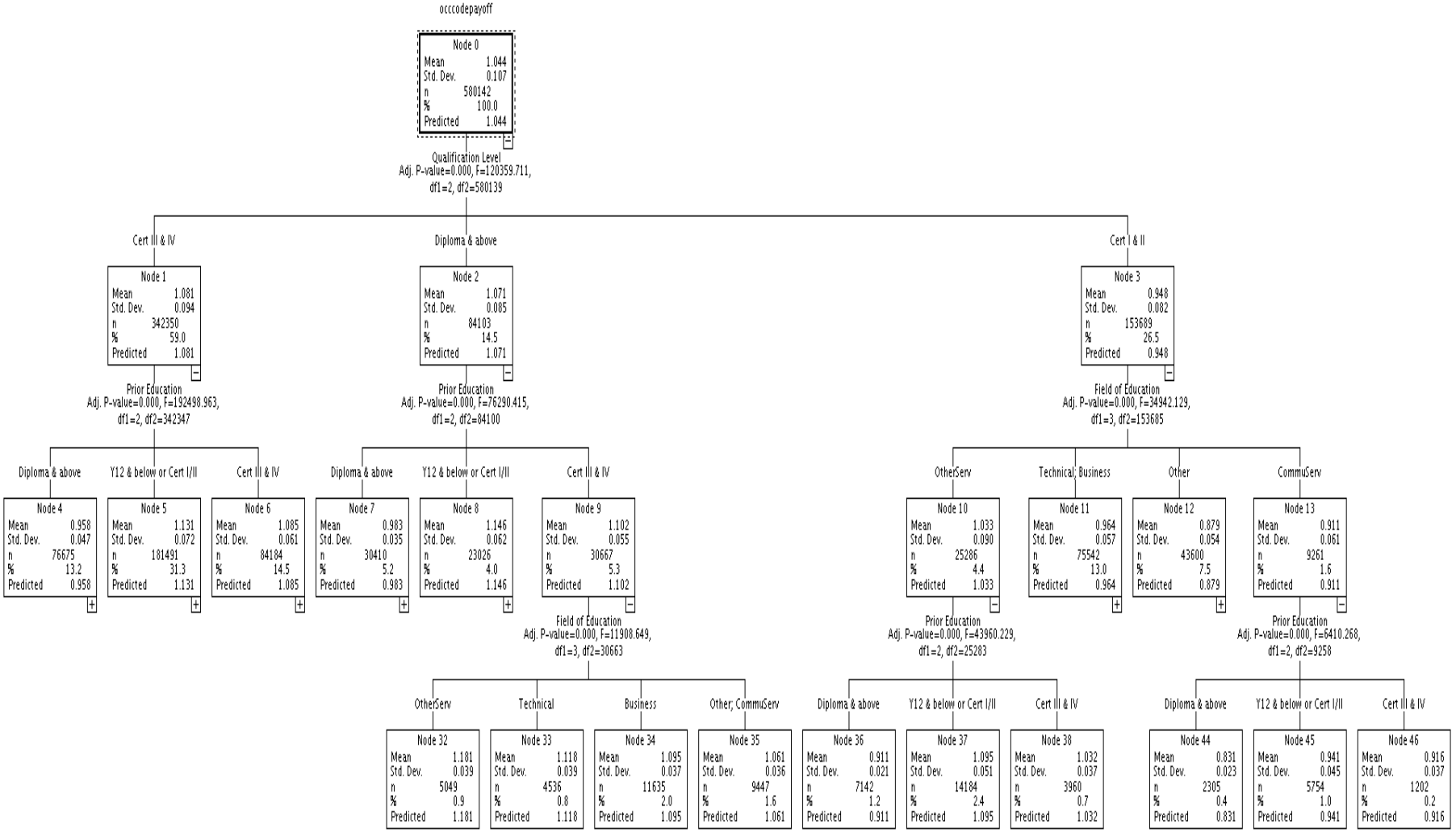

Figure 4.8 CHAID diagram – Completion benefit to 'Occupational status after training'

Examining the CHAID chart in figure 4.8 revealed that the qualification levels of Certificate I and II had a negative pay-off to completion, although there were some student groups within these categories that gained a minor positive pay-off, such as those who had a prior education below diploma in the 'Other services' field of education. This overall lack of benefit from completion for lower-level VET qualifications in respect to occupational status mirrored the similar pattern in the salary pay-off analysis. This appeared to be a logical association, as it can be assumed that occupational status and personal income[6] are intrinsically linked. The benefit of completion appeared to be largest for those completing a qualification level above Certificate II and a prior education below Certificate III (1.14 or 14% enrolled in Diploma and above, and 1.13 or 13% enrolled in Certificate II and IV).

The benefit of completion with respect to student satisfaction

Student satisfaction with their vocational education is generally quite high. In the 2011 student outcome survey 89.3% of students indicated that they were either satisfied or very satisfied with their training (NCVER 2011b). To determine if completion was also a substantial factor in student satisfaction we used the five-point Likert scale response to the SOS question: How would you rate, on average, your satisfaction with the overall quality of the training? Overall, I was satisfied with the quality of this training: to which the responses 'strongly agree', 'agree', 'neither agree nor disagree', 'disagree', and 'strongly disagree' were available (NCVER 2011e). These responses were then converted into an ordinal variable with values from one to five. While not strictly continuous, we used this variable as the dependent variable for subsequent OLS regression in order to predict satisfaction scores for completers and non-completers, applying our standard predictors. The model estimates were presented in table 4.9.

[6] Occupational status and annual income were significantly correlated in the 2011 SOS, albeit not strongly ($r=0.36$, $p<0.01$).

Table 4.9 Student satisfaction regression models

Satisfaction with training		Completers			Non-completers		
		B	Sig.	Beta	B	Sig.	Beta
Sex	Female	-0.02	<0.01	-0.01	-0.04	<0.01	-0.02
	Male	0.00			0.00		
SEIFA	Most disadvantaged	0.10	<0.01	0.04	0.01	0.18	0.00
	Very disadvantaged	0.08	<0.01	0.04	0.04	<0.01	0.02
	Somewhat disadvantaged	0.05	<0.01	0.02	-0.02	<0.01	-0.01
	Little disadvantaged	0.03	<0.01	0.02	0.01	0.153	0.00
	Least disadvantaged	0.00			0.00		
Field of education	Technical	-0.11	<0.01	-0.05	-0.03	<0.01	-0.01
	Business	-0.03	<0.01	-0.02	-0.14	<0.01	-0.06
	Comm. services	0.01	0.066	0.01	-0.03	<0.01	-0.01
	Other services	-0.05	<0.01	-0.02	0.04	<0.01	0.02
	Other	0.00			0.00		
Remoteness	City	-0.03	<0.01	-0.02	-0.19	<0.01	-0.10
	Regional	-0.02	<0.01	-0.01	-0.08	<0.01	-0.04
	Remote	0.00			0.00		
Age group	Age 15–24	0.02	<0.01	0.01	-0.10	<0.01	-0.05
	Age 25–44	-0.01	0.133	0.00	-0.01	0.06	-0.01
	Age 45+	0.00			0.00		
Study mode	Part-time	-0.07	<0.01	-0.04	0.03	0.084	0.00
	Full-time	0.00			0.00		
Status before training	Employed	0.01	0.087	0.00	0.13	<0.01	0.06
	Unemployed	0.03	0	0.01	0.08	<0.01	0.03
	Not in labour force	0.00			0.00		
Prior education	Diploma and above	-0.11	<0.01	-0.05	-0.06	<0.01	-0.03
	Cert III/IV	-0.01	<0.01	-0.01	-0.05	<0.01	-0.02
	Cert I/II	0.00			0.00		
Qualification level	Diploma and above	-0.10	<0.01	-0.05	-0.30	<0.01	-0.09
	Cert III/IV	-0.08	<0.01	-0.04	-0.16	<0.01	-0.09
	Cert I/II	0.00			0.00		
	Constant	4.40	<0.01		4.36	<0.01	
		Adj. rsq		0.01	Adj rsq		0.03
		F = 194.5 (p<0.01)			F = 345.8 (p<0.01)		

It became quite apparent that the fit of these two models was fairly moderate. While there was some overall significance and there were several significant predictors, it was clear that the overall amount of explained variance was small (1% for completers and 3% for non-completers). Generally it was concluded that in addition to the variables included in the model, there were other, unobserved, traits that better predicted student satisfaction. However, as the SOS did not collect other data that could have provided more utility in predicting student satisfaction we used the above models to derive

estimates of individual benefits from completion. The overall benefit from completion in terms of student satisfaction was calculated as 1.042, indicating a modest overall benefit. The distribution of benefit has been displayed in the histogram in figure 4.9.

Figure 4.9 Histogram of completion benefit for student satisfaction

Overall, a large majority of students (83.6%; table 4.10) benefited from completion with higher satisfaction with their training, while 16.4% of students received a negative payoff for this variable.

Table 4.10 Payoff summary 'Satisfaction with the training'

Satisfaction with the training	Number	Per cent
Payoff > 1	483,648	83.6
Payoff <= 1	95,171	16.4

The CHAID diagram (figure 4.10) suggested that the most significant split was between full-time and part-time students. However, the difference between these two student categories was relatively modest, as full-time students (1.064) drew only a slightly greater benefit than part-time students (1.027) from completing their program. Differences in completion benefits for various

categories of students were small throughout the CHAID diagram, which may have been partly due to the only modest predictive power of completion and non-completion models.

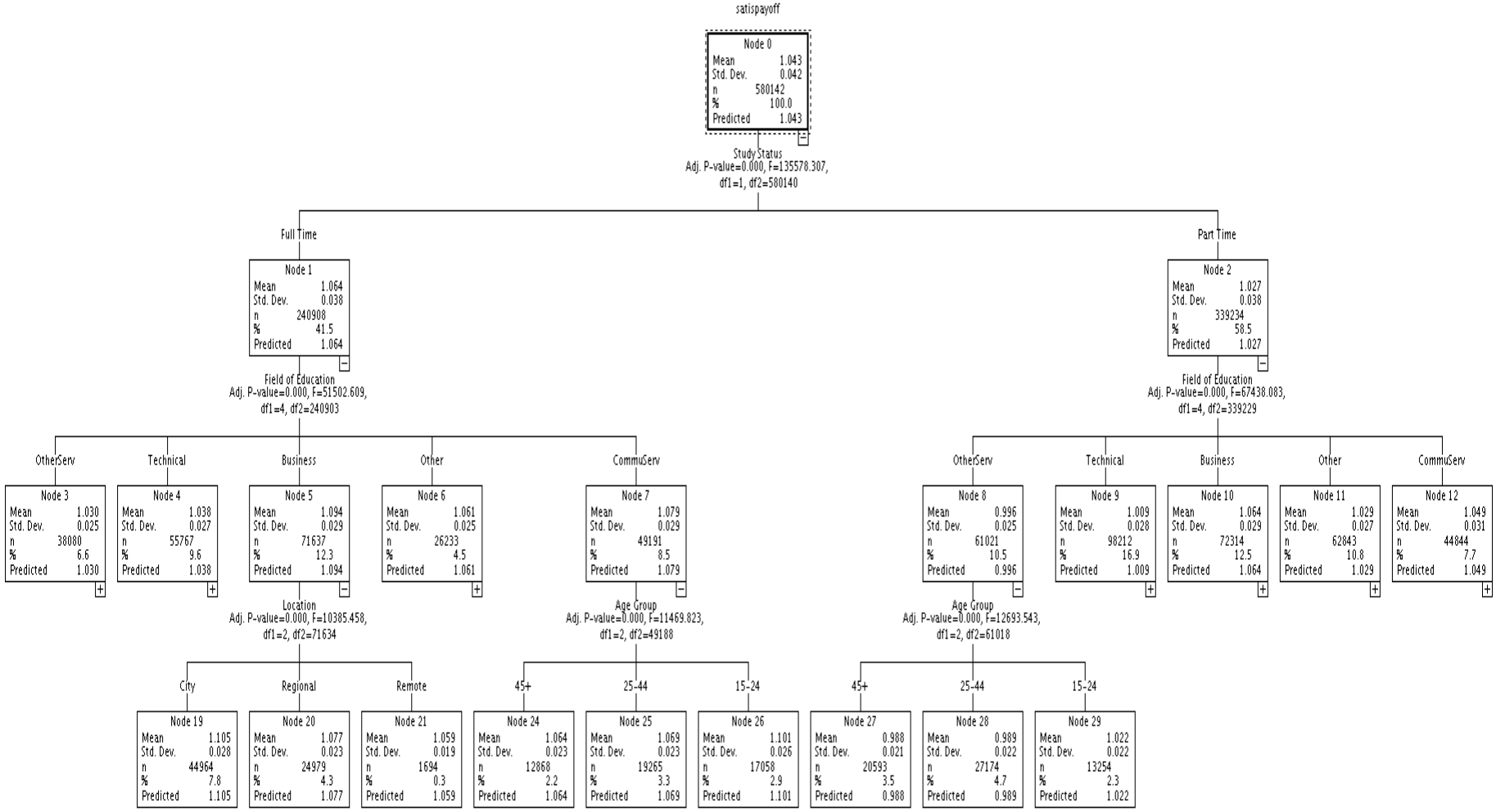

Figure 4.10 CHAID diagram – Completion benefit to 'Satisfaction with the training'

The benefit of completion with respect to further studies after training

The benefit of completion with respect to pursuing further study made for an interesting analysis subject. We hypothesised significant benefits for those who complete as it seemed reasonable to assume that the completion of one qualification may have been a prerequisite for enrolment into a host of subsequent qualifications. Furthermore, in previous sections we speculated that the absence of direct material benefits for lower-level VET qualification may be due to the fact that such qualifications can be used as stepping stones into higher-level qualifications. Consequently, we expected a particularly worthwhile benefit with respect to pursuing further study for those students who completed lower-level qualifications. The dependent variable in completion and non-completion models was the dichotomous indicator that further study was undertaken after the initial training, along with the predictor variables used in previous models. Estimates for both logistic regression models can be found in table 4.11.

Table 4.11 Further study after training logistic regression models

Further study after training		Completers			Non-completers		
		B	Sig.	Exp(B)	B	Sig.	Exp(B)
Sex	Female	0.12	<0.01	1.13	-0.01	0.40	0.99
	Male	0.00			0.00		
SEIFA	Most disadvantaged	-0.01	0.37	0.99	-0.08	<0.01	0.93
	Very disadvantaged	0.01	0.62	1.01	-0.17	<0.01	0.85
	Somewhat disadvantaged	0.03	0.03	1.03	-0.25	<0.01	0.78
	Little disadvantaged	0.01	0.36	1.01	-0.06	<0.01	0.95
	Least disadvantaged	0.00			0.00		
Field of education	Technical	-0.26	<0.01	0.77	-0.24	<0.01	0.78
	Business	0.16	<0.01	1.18	-0.38	<0.01	0.69
	Comm. services	0.29	<0.01	1.34	-0.04	0.04	0.96
	Other services	0.11	<0.01	1.12	-0.16	<0.01	0.85
	Other	0.00			0.00		
Remoteness	City	0.09	<0.01	1.09	0.16	<0.01	1.18
	Regional	-0.05	0.02	0.95	0.16	<0.01	1.18
	Remote	0.00			0.00		
Age group	Age 15–24	0.71	<0.01	2.03	0.94	<0.01	2.55
	Age 25–44	0.28	<0.01	1.32	0.35	<0.01	1.42
	Age 45+	0.00			0.00		
Study mode	Part-time	-0.24	<0.01	0.79	-0.04	0.36	0.96
	Full-time	0.00			0.00		
Status before training	Employed	-0.31	<0.01	0.73	0.00	0.86	1.00
	Unemployed	-0.23	<0.01	0.79	0.02	0.33	1.02
	Not in labour force	0.00			0.00		
Prior education	Diploma and above	0.09	<0.01	1.10	0.27	<0.01	1.31
	Cert III/IV	0.10	<0.01	1.11	0.07	<0.01	1.08
	Cert I/II				0.00		
Qualification level	Diploma and above	-0.36	<0.01	0.70	0.33	<0.01	1.39
	Cert III/IV	-0.43	<0.01	0.65	0.03	0.01	1.04
	Cert I/II	0.00			0.00		
	Constant	-0.612	<0.01	0.54	-2.22	<0.01	0.11
		Pseudo rsq		0.05	Pseudo rsq		0.04
		Chi Sq = 12586.9		(p<0.01)	Chi Sq = 4982.1		(p<0.01)

The two models explained the variance in completers and non-completers only moderately well, as evidenced by a relatively low pseudo r-squared of 0.05 and 0.04 respectively. Female sex was associated with a 13% higher likelihood than males for further study in completers, while sex played no differentiating role in determining further study in non-completers. Among non-completers, less advantaged students were also less likely to engage in further studies, whereas among completers socioeconomic status played no such role. Students from remote areas were less likely to engage in further studies than their

urban counter parts, irrespective of whether they were completers or not. The most significant differentiator in respect to further study was student age: students less than 25 years old were more than twice as likely than 45+ year-olds to enrol in further study, regardless if they were completers or non-completers. The two models could then be used to predict probabilities for further study. The resulting overall mean was 2.25 and the associated histogram can be seen in figure 4.11.

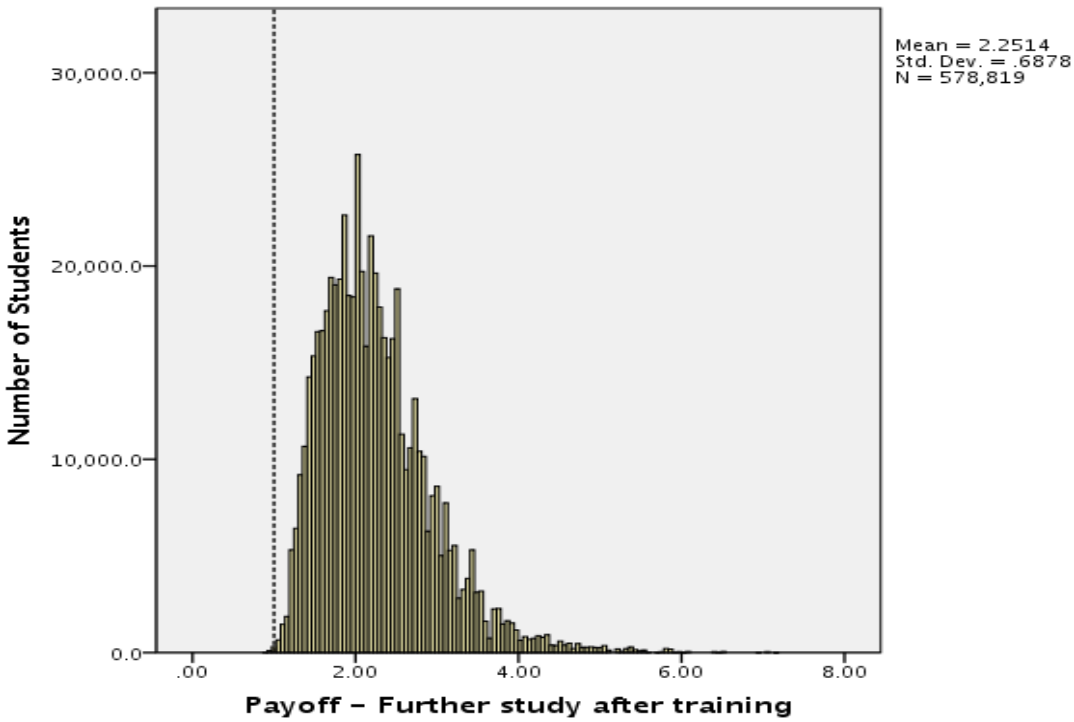

Figure 4.11 Histogram of completion benefit for further study after training

The mean and the histogram signal a significant departure from the outcomes of previous analyses. The overall mean of 2.25 denoted a very strong benefit of completion when engaging in further study. This result indicated that the probability of a completer enrolling in further study was more than twice as high as the probability of a non-completer. It was also evident that practically everyone (99.9%; table 4.12) gains some benefit from completion with respect to pursuing further study. This was in contrast to the previous analyses in this study, where there was always a significant proportion of students who did not receive a positive benefit from completion.

Table 4.12 Payoff summary 'Further study after training

Enrolment in further study after training	Number	Per cent
Payoff > 1	578,556	99.9
Payoff <= 1	263	0.1

Evaluating the CHAID diagram presented in figure 4.12 we noted that the gains from completion were strongest for those enrolled in lower-level qualifications, for example, Certificates I and II, who were 175% more likely to enrol in further study if they completed their initial VET qualification. Completers of higher VET qualifications, for example, Certificates III and IV, were 111% more likely than their non-completing counterparts to enrol in further study. Those who completed a diploma were 89% more likely to take up further study than those who did not complete their diploma qualification. Nested below the Certificate I/II category were student groups who exhibit a substantially higher propensity for further study than the group mean, provided they complete. For instance, students who completed their certificate I or II in a business discipline were more than 280% more likely to enrol in further study than non-completers in the same category.

The analysis of the benefit of completion in terms of further study showed that completion yielded strong and universal benefits that outweighed completion benefits in other categories. The results of this analysis also appeared to confirm our earlier supposition that completing lower-level VET qualification could in many cases be considered as a stepping stone into further education.

furtherstudpayoff

Node 0
Mean 2.252
Std. Dev. 0.688
n 580142
% 100.0
Predicted 2.252

Qualification Level
Adj. P-value=0.000, F=74262.700,
df1=2, df2=580139

Cert III & IV

Node 1
Mean 2.115
Std. Dev. 0.575
n 342350
% 59.0
Predicted 2.115

Field of Education
Adj. P-value=0.000, F=91481.610,
df1=4, df2=342345

OtherServ

Node 4
Mean 2.174
Std. Dev. 0.384
n 61585
% 10.6
Predicted 2.174

Technical

Node 5
Mean 1.663
Std. Dev. 0.285
n 98998
% 17.1
Predicted 1.663

Business

Node 6
Mean 2.724
Std. Dev. 0.523
n 75816
% 13.1
Predicted 2.724

Age Group
Adj. P-value=0.000, F=18935.777,
df1=2, df2=75813

45+

Node 23
Mean 3.177
Std. Dev. 0.532
n 15480
% 2.7
Predicted 3.177

25-44

Node 24
Mean 2.823
Std. Dev. 0.463
n 30642
% 5.3
Predicted 2.823

15-24

Node 25
Mean 2.384
Std. Dev. 0.313
n 29694
% 5.1
Predicted 2.384

Other

Node 7
Mean 1.701
Std. Dev. 0.320
n 43240
% 7.5
Predicted 1.701

CommuServ

Node 8
Mean 2.318
Std. Dev. 0.447
n 62711
% 10.8
Predicted 2.318

Age Group
Adj. P-value=0.000, F=16908.079,
df1=2, df2=62708

45+

Node 29
Mean 2.649
Std. Dev. 0.435
n 17447
% 3.0
Predicted 2.649

25-44

Node 30
Mean 2.377
Std. Dev. 0.384
n 23892
% 4.1
Predicted 2.377

15-24

Node 31
Mean 1.984
Std. Dev. 0.247
n 21372
% 3.7
Predicted 1.984

Diploma & above

Node 2
Mean 1.893
Std. Dev. 0.444
n 84103
% 14.5
Predicted 1.893

Cert I & II

Node 3
Mean 2.754
Std. Dev. 0.760
n 153689
% 26.5
Predicted 2.754

Field of Education
Adj. P-value=0.000, F=30482.027,
df1=4, df2=153684

OtherServ

Node 14
Mean 2.814
Std. Dev. 0.518
n 25286
% 4.4
Predicted 2.814

Age Group
Adj. P-value=0.000, F=4827.700,
df1=2, df2=25283

45+

Node 44
Mean 3.206
Std. Dev. 0.529
n 7013
% 1.2
Predicted 3.206

25-44

Node 45
Mean 2.809
Std. Dev. 0.460
n 9130
% 1.6
Predicted 2.809

15-24

Node 46
Mean 2.519
Std. Dev. 0.332
n 9143
% 1.6
Predicted 2.519

Technical

Node 15
Mean 2.342
Std. Dev. 0.414
n 41599
% 7.2
Predicted 2.342

Business

Node 16
Mean 3.617
Std. Dev. 0.799
n 33943
% 5.9
Predicted 3.617

Age Group
Adj. P-value=0.000, F=23193.494,
df1=2, df2=33940

45+

Node 50
Mean 4.614
Std. Dev. 0.753
n 6569
% 1.1
Predicted 4.614

25-44

Node 51
Mean 4.036
Std. Dev. 0.626
n 7402
% 1.3
Predicted 4.036

15-24

Node 52
Mean 3.133
Std. Dev. 0.355
n 19972
% 3.4
Predicted 3.133

Other

Node 17
Mean 2.379
Std. Dev. 0.481
n 43600
% 7.5
Predicted 2.379

CommuServ

Node 18
Mean 3.041
Std. Dev. 0.650
n 9261
% 1.6
Predicted 3.041

Figure 4.12 CHAID diagram – Completion benefit to 'Further study after training'

The benefit of completion with respect to improved employment conditions?

The final benefit of completion outcome we wanted to examine in this study was the 'improved employment conditions'. This outcome is a variation of the 'employed after training' outcome and was derived from three indicators in the student outcome survey. For this outcome indicator, an affirmative response to any of the three questions 'Were you not employed before training, but employed after?', 'Are you employed in a higher skill level after training?', or 'Was there at least one job-related benefit?' was defined as improved employment condition. This yielded a dichotomous indicator and therefore logistic regression was performed to estimate completion and non-completion models. Independent variables used as predictors were the same as in the previous models. Estimated coefficients and model statistics for both models can be found in table 4.13.

Table 4.13 Improved employment conditions logistic regression models

Improved employment conditions after training		Completers			Non-completers		
		B	Sig.	Exp(B)	B	Sig.	Exp(B)
Sex	Female	-0.09	<0.01	0.91	-0.18	<0.01	0.84
	Male	0.00			0.00		
SEIFA	Most disadvantaged	-0.28	<0.01	0.76	-0.20	<0.01	0.82
	Very disadvantaged	-0.16	<0.01	0.85	-0.18	<0.01	0.83
	Somewhat disadvantaged	-0.11	<0.01	0.90	-0.06	<0.01	0.94
	Little disadvantaged	-0.03	0.01	0.97	-0.04	0.01	0.97
	Least disadvantaged	0.00			0.00		
Field of education	Technical	0.68	<0.01	1.98	0.21	<0.01	1.23
	Business	0.11	<0.01	1.12	-0.16	<0.01	0.85
	Comm. services	0.18	<0.01	1.20	-0.28	<0.01	0.76
	Other services	0.41	<0.01	1.51	0.09	<0.01	1.10
	Other	0.00			0.00		
Remoteness	City	-0.38	<0.01	0.68	-0.14	<0.01	0.87
	Regional	-0.18	<0.01	0.83	-0.03	0.06	0.97
	Remote	0.00			0.00		
Age group	Age 15–24	0.36	<0.01	1.43	0.18	<0.01	1.20
	Age 25–44	0.18	<0.01	1.19	0.22	<0.01	1.25
	Age 45+	0.00			0.00		
Study mode	Part-time	0.17	<0.01	1.19	-0.07	0.05	0.93
	Full-time	0.00			0.00		
Status before training	Employed	0.92	<0.01	2.52	0.98	<0.01	2.65
	Unemployed	0.25	<0.01	1.29	0.59	<0.01	1.81
	Not in labour force	0.00			0.00		
Prior education	Diploma and above	-0.13	<0.01	0.88	0.08	<0.01	1.09
	Cert III/IV	-0.02	0.02	0.98	0.17	<0.01	1.19
	Cert I/II	0.00			0.00		
Qual level	Diploma and above	0.53	<0.01	1.69	0.38	<0.01	1.46
	Cert III/IV	0.60	<0.01	1.82	0.38	<0.01	1.47
	Cert I/II	0.00			0.00		
	Constant	-0.67	<0.01	0.51	-0.89	<0.01	0.41
		Pseudo rsq		0.10	Pseudo rsq		0.08
		Chi Sq = 25726.8		(p<0.01)	Chi Sq = 15334.4		(p<0.01)

Pseudo r-squared was slightly higher than in the 'further study' models with 0.10 for graduates and 0.08 for module completers. For completers and non-completers alike, females were slightly less likely than males to experience improved employment conditions after training. Lower socioeconomic backgrounds tended to be associated with a smaller propensity of improved employment conditions in completers and non-completers. As in previously examined outcomes, remote students fared better in improving their

employment circumstances than regional and urban students. This was the case for completers and non-completers. Interestingly, students with prior VET qualifications higher than certificate II were less likely than those with lower prior qualifications to improve their employment condition among completers, while those who did not complete their qualification and had higher prior VET qualifications were more likely than those with lower prior qualifications to experience improved employment conditions.

After predicting improved employment conditions probabilities for completer and non-completer models we could calculate the overall mean of the completion benefit as 1.247, indicating an approximate 25% overall benefit of completion. The distribution of the employment benefit can be seen in figure 4.13.

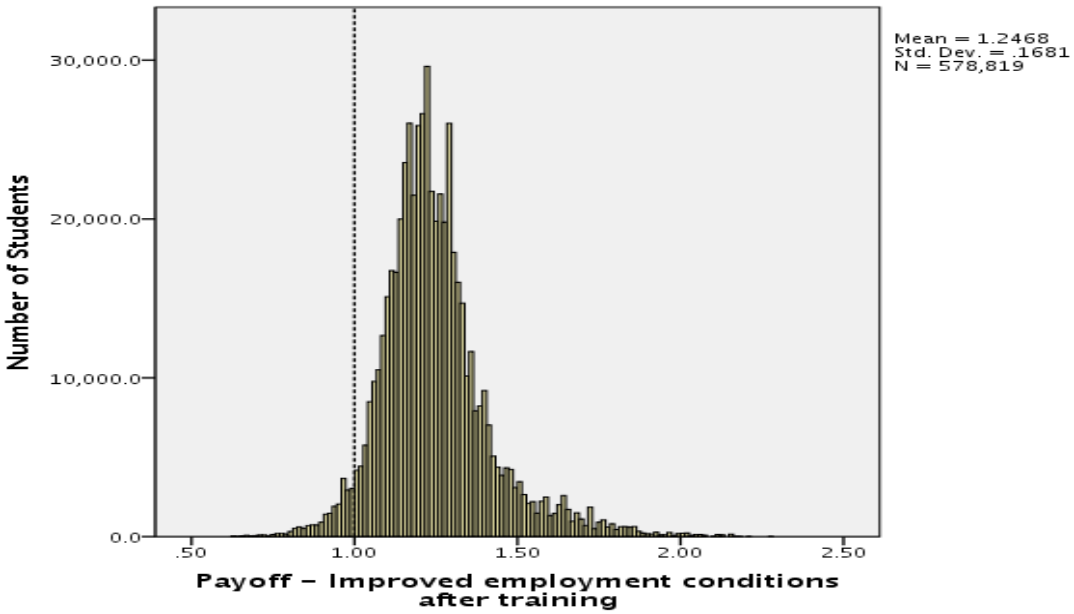

Figure 4.13 Histogram of completion benefit for improved employment conditions

The histogram revealed that almost all students (96.1%; table 4.14) gained a benefit from completion in the form of improved employment conditions.

Table 4.14 Payoff summary 'Improved employment conditions after training'

Improved employment conditions after training	Number	Per cent
Payoff > 1	556,120	96.1
Payoff <= 1	22,699	3.9

As in previous analyses we expected that the benefit was quite unevenly distributed across different groups of students and performed a CHAID analysis (figure 4.14). Surprisingly, previously unemployed students drew the smallest pay-off (14%) from completion, while those not previously in the labour force gained the strongest benefit (48%). As those not previously in the labour force will have been predominantly school leavers, this indicated that completion of their qualification was particularly worthwhile for this demographic. Among all three pre-training employment categories, part-time study mode imparted a stronger benefit of completion than full-time study. In this context it was quite likely that part-time students were also part-time employees and that the completion of a qualification led to a direct reward of improved employment conditions, such as a higher salary and/or a promotion. Of those students who were unemployed before their training and who studied full time, enrolling in technical and community service areas yielded the greatest benefit (13 and 19 % respectively).

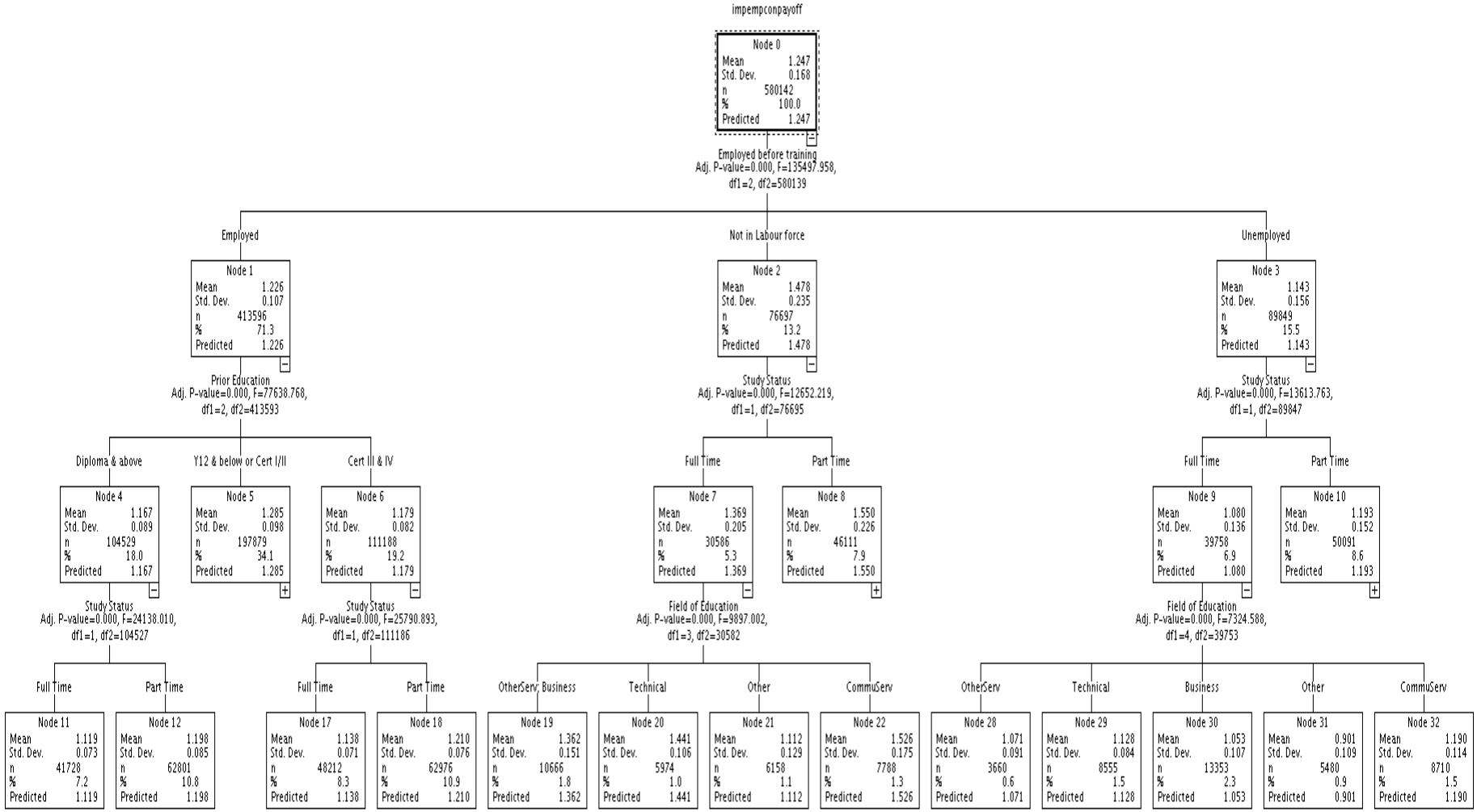

Figure 4.14 CHAID diagram – Completion benefit to 'Improved employment conditions after training'

4.1.2.2 Concluding thoughts about the benefit of completion

Previous studies on the benefit of completing a VET qualification have found that the returns to completion were often modest (for example, Preston 1997; Ryan, 2002; Karmel & Nguyen, 2006; Stanwick, 2005). This appeared to be particularly so for lower qualifications (Karmel & Nguyen, 2006) and where tangible, material outcomes such as salaries were concerned (Preston, 1997; Ryan, 2002). In our analysis here we presented a methodology that facilitated the allocation of predicted scores for various training-related outcomes for each individual student based on two scenarios: First, that the student completed the commenced training, and second, that the student discontinued the training before completing the qualification. This methodology enabled the estimation of a predicted outcome score for both of the two scenarios for each individual student, which in turn allowed us to calculate a benefit of completion for every student. The advantage of this methodology was that after the individual benefits to completion had been established, the analysis could then proceed to identify and group students into practically every possible combination of student characteristics and then establish potential completion benefits for this group. We chose Chi Squared Automated Interaction Detection to analyse these groups, as this type of analysis was useful to not only identify groups that significantly differ, but also allowed conclusions about the magnitude of differences between student groups that share certain characteristics and the relation to the outcome variable in question.

We found that completion of a qualification was mostly beneficial in terms of the outcomes we evaluated, although there was significant variation in pay-offs between some student groups. Among the pay-offs were some rather surprising results, such as the overall lack of completion benefit in terms of post-training salaries, and the relatively small benefit for previously unemployed students in terms of employment outcomes. Other results of this

study were that we have been able to confirm the relative lack of benefits from completion for lower-level VET qualifications that previous research had uncovered (Ryan, 2002; Karmel & Nguyen, 2006), with the exception of the overwhelming benefit of completion for this group with respect to further study. We compiled the mean benefits for the main student groups in one table (table 4.15) to allow a glance at the benefits for individual groups and all the training outcome variables we have examined. It became immediately obvious that the main completion benefit was derived from the further study outcome. Students of almost all educational and demographic background characteristics were more than twice as likely to enrol in further study if they completed their initial training. Other noteworthy themes were the overall lack of benefit of completion in terms of salary and the only modest benefit to completion in terms of occupational status and satisfaction.

Table 4.15 Overview of benefits to completion for different groups of students

Category		Payoffs					
		Employ-ment	Full-time salary	Occup. status	Satis-faction	Further study	Improved employ-ment condition
Sex	Female	1.18	0.98	1.01	1.05	2.47	1.27
	Male	1.13	0.99	1.08	1.03	2.04	1.22
Age group	15–24	1.25	0.99	1.06	1.06	2.02	1.32
	25–44	1.07	0.98	1.04	1.03	2.28	1.17
	45+	1.14	0.98	1.03	1.03	2.61	1.24
Remoteness	City	1.16	0.99	1.04	1.06	2.25	1.22
	Regional	1.15	0.98	1.05	1.03	2.22	1.27
	Remote	1.11	0.98	1.06	1.00	2.59	1.32
SEIFA	Most disadvantaged	1.14	1.01	1.07	1.05	2.19	1.23
	Very disadvantaged	1.19	0.99	1.05	1.04	2.36	1.29
	Somewhat disadvantaged	1.14	1.00	1.07	1.05	2.50	1.23
	Little disadvantaged	1.13	0.95	1.02	1.04	2.10	1.24
	Least disadvantaged	1.16	0.97	1.01	1.04	2.00	1.22
Qualification level	Cert I & II	1.17	0.88	0.95	1.02	2.75	1.24
	Cert III & IV	1.16	1.02	1.08	1.04	2.11	1.26
	Diploma & above	1.09	1.04	1.07	1.08	1.89	1.22
Field of education	Business	1.17	1.01	1.04	1.08	2.81	1.24
	Community service	1.22	0.95	1.03	1.06	2.28	1.37
	Other	1.14	0.95	0.95	1.04	2.03	1.13
	Other services	1.09	1.03	1.11	1.01	2.29	1.22
	Technical	1.16	0.98	1.07	1.02	1.82	1.26
Labour force status before training	Employed	1.03	0.99	1.04	1.03	2.14	1.23
	Unemployed	1.27	0.94	1.04	1.06	2.36	1.14
	Not in labour force	1.67	1.02	1.06	1.07	2.70	1.48
Study status	Part-time	1.11	0.98	1.03	1.03	2.14	1.29
	Full-time	1.21	0.99	1.07	1.06	2.41	1.19
Prior education	Diploma & above	1.05	0.96	0.94	1.03	2.07	1.16
	Cert III & IV	1.06	0.96	1.07	1.05	2.30	1.18
	Y12 & below or Cert I/II	1.24	1.00	1.08	1.05	2.31	1.31
Total	**Mean**	**1.16**	**0.98**	**1.04**	**1.04**	**2.25**	**1.25**

4.1.3 Is the way to completion paved with good intentions?

In the previous sections we demonstrated the extent to which students with similar characteristics benefited from completion with respect to various outcomes. It seems not unreasonable to assume that many students will consider the rewards of completion when they decide to complete or discontinue their training. Similarly, one could speculate that an increasing benefit of completion should lead to increased completion numbers. It is thus of interest to examine whether there is a link between completion and the benefit thereof. Furthermore, we were interested in determining if, and if so, how strong any association between intended and actual completion is.

The student outcome survey canvassed students' attitudes on whether they achieved their main reason for undertaking their training, which was answered in the affirmative by a surprisingly large percentage of respondents (82%) despite the overall low completion rate. This high percentage appeared to support the theory that many students enrolled in their training with the intention to gain specific skills rather than a qualification. In order to gain a deeper understanding of students' intention to complete at the time of, or shortly after, their enrolment, NCVER conducted the Student Intentions Survey in 2011. The surprising result was that an overwhelming majority of students (93%) did indeed intend to complete their qualification.

In this second part of our portfolio paper we intended to shine a light on issues surrounding the likelihood of completion, the gap between the intention to complete and actual completion and whether the benefits of completion we have determined in the first part of this paper have any impact on completion.

4.1.3.1 Data and methods

In this part of the paper we made use of two surveys conducted in 2011: The SOS and the Student Intentions Survey (SIS). We chose the 2011 waves of both surveys as 2011 was the only occasion when the SIS was conducted, which

coincided with a 'large' SOS year. While the 2011 SOS consisted of approximately 110,000 completed responses in order to enable reportable data for every Australian TAFE institute, the SIS featured a much smaller received response of 10,660 out of a sample of 23,086 students, as this survey was designed to provide statistically meaningful estimates at the national level only (NCVER 2011c). While both surveys gathered a similar level of administrative demographic and educational student background data, the SIS limited the information that was collected to issues surrounding the willingness to complete, knowledge about the training at enrolment and attitudes toward the training and vocational education.

In this analysis we were interested specifically in how completion related to the intention to complete and also how the benefit to completion related to completion. The Student Intentions Survey derived the intention to complete via the response to this survey question: 'Is it your intention to A) complete the training required to gain the qualification for the [course name] or is it your intention to B) complete some subjects of the [course name], rather than complete the whole course?'. As the samples for SIS and SOS were drawn from two different populations and thus contained different individual students, we had to devise a method which would allow us to use the information about the intention to complete that we gained from the SIS to the SOS which contained the information about actual completion. To do this we first estimated the intention to complete from the student intentions survey, using the variable described above as the dependent variable. When doing this we were mindful of only using variables as predictors that had an equivalent in the student outcome survey. The result of this logistic regression can be found in table 4.16. It is worth noting the females were 38% more likely to intend to complete than males, and younger age groups were more likely to intend to complete than older age groups. Those who were enrolled full-time were three times more likely to intend to complete than part-time students, and non-

remote students had a greater propensity to intend to complete than remote students. Not surprisingly, those studying with employment-related goals were more likely to intend to complete than those with personal reasons for study.

Table 4.16 Logistic regression parameters for intention to complete

		Probability of intending to complete				
		B	S.E.	Wald	Sig.	Exp(B)
Prior education	<Y12 & Certificate I/II	0.01	0.04	0.0	0.85	1.01
Prior education	Certificate III/IV	-0.13	0.04	11.9	<.01	0.88
Prior education	Diploma and above	-0.21	0.04	29.9	<.01	0.81
Prior education	Y12	0.00				
Field of education	Business	-0.02	0.04	0.3	0.58	0.98
Field of education	Community services	0.39	0.05	69.0	<.01	1.48
Field of education	Other	-1.09	0.05	568.0	<.01	0.34
Field of education	Other services	-0.72	0.04	339.1	<.01	0.49
Field of education	Technical	0.00				
Sex	Female	0.32	0.03	130.1	<.01	1.38
Sex	Male	0.00				
Age group	25–34 years	0.34	0.04	64.3	<.01	1.40
Age group	35–45 years	-0.05	0.04	1.8	0.18	0.95
Age group	<25 years	0.60	0.04	237.9	<.01	1.82
Age group	>45 years	0.00				
Qualification level	Certificate I/II	-0.73	0.07	109.6	<.01	0.48
Qualification level	Certificate III/IV	-0.16	0.06	6.1	0.01	0.85
Qualification level	Diploma	-0.52	0.07	59.7	<.01	0.59
Qualification level	Other	0.00				
Study status	Full-time	1.12	0.03	1,368.1	<.01	3.05
Study status	Part-time	0.00				
Employment status	Employed	0.00	0.03	0.0	0.94	1.00
Employment status	Not employed	0.00				
Remoteness	City	0.25	0.09	7.2	<.01	1.28
Remoteness	Regional	0.27	0.09	8.6	<.01	1.31
Remoteness	Remote	0.00				
SEIFA	Least disadvantaged	0.06	0.04	2.0	0.16	1.06
SEIFA	Somewhat disadvantaged	0.15	0.04	12.7	<.01	1.17
SEIFA	Most disadvantaged	-0.14	0.05	9.0	<.01	0.87
SEIFA	Little disadvantaged	-0.38	0.04	87.4	<.01	0.69
SEIFA	Very disadvantaged	0.00				
Reason for study	Employment related	0.47	0.03	250.3	<.01	1.60
Reason for study	Further study related	-0.28	0.05	35.9	<.01	0.75
Reason for study	Personal	0.00				
Constant		2.10	0.12	282.7	<.01	
N = 6971 Chi-Squared = 5864.7 R-squared(pseudo) = 0.593						

Subsequently we used the coefficient estimates from the intention to complete model in SIS model shown in table 4.16 to perform an out-of-sample

prediction using the very same predictor variables from the SOS to estimate the intention to complete probabilities for every individual in the SOS.

In order to develop a continuous variable that encompassed the actual probability of completing we then utilised the SOS and its completion variable (for example, graduate vs module completer) as dependent variable and developed a similar model which was then used to predict actual completion. The model parameters can be found in table 4.17.

It was evident from the 'intention to complete' and 'actual completion' models that the estimated coefficients were reasonably similar, albeit mostly more pronounced in the actual completion model. This could be seen in the much larger propensity of diploma and above students to complete in relation to other qualification levels compared with the intention model. Similarly, in the actual completion model full-time students were 15 times more likely to complete than their part-time counterparts, whereas in the intention model they were only three times as likely.

Table 4.17 Logistic regression parameters for actual completion

		Probability of actual completion				
		B	S.E.	Wald	Sig.	Exp(B)
Prior education	<Y12 & Certificate I/II	0.12	0.02	25.90	<.01	1.13
Prior education	Certificate III/IV	-0.05	0.02	4.24	0.04	0.95
Prior education	Diploma & above	-0.29	0.03	133.32	<.0001	0.74
Prior education	Y12	0.00				
Field of education	Business	0.35	0.03	165.73	<.01	1.42
Field of education	Community services	0.55	0.03	410.76	<.01	1.74
Field of education	Other	-0.11	0.05	4.61	0.03	0.89
Field of education	Other services	-0.83	0.03	677.87	<.01	0.43
Field of education	Technical	0.00				
Sex	Female	0.14	0.02	86.81	<.01	1.15
Sex	Male	0.00				
Age group	25–34 years	-0.04	0.03	2.36	0.12	0.96
Age group	35–45 years	-0.07	0.03	8.94	0.00	0.93
Age group	<25 years	0.53	0.02	494.38	<.01	1.70
Age group	>45 years	0.00				
Qualification level	Cert I/II	1.07	0.09	128.18	<.01	2.92
Qualification level	Certificate III/IV	1.75	0.09	352.30	<.01	5.77
Qualification level	Diploma	2.32	0.10	579.65	<.01	10.14
Qualification level	Other	0.00				
Study status	Full-time	2.71	0.03	7,725.67	<.01	15.07
Study status	Part-time	0.00				
Employment status	Employed	0.10	0.02	35.49	<.01	1.10
Employment status	Not employed	0.00				
Remoteness	City	0.19	0.02	62.75	<.01	1.21
Remoteness	Regional	0.03	0.02	1.20	0.27	1.03
Remoteness	Remote	0.00				
SEIFA	Least disadvantaged	0.01	0.03	0.13	0.71	1.01
SEIFA	Somewhat disadvantaged	0.00	0.03	0.00	0.98	1.00
SEIFA	Most disadvantaged	0.01	0.03	0.20	0.66	1.01
SEIFA	Little disadvantaged	-0.02	0.02	0.89	0.35	0.98
SEIFA	Very disadvantaged	0.00				
Reason for study	Employment	0.08	0.03	6.95	0.01	1.08
Reason for study	Further	0.02	0.05	0.10	0.75	1.02
Reason for study	Personal	0.00				
Constant		0.05	0.10	0.30	0.58	
	N = 67,502 Chi-Squared = 55,773.4 R-squared(pseudo) = 0.761					

4.1.3.2 Probability of completing versus probability of intention to complete

After preparing probability scores for completion and intention to complete for each individual student in the SOS and removing students who were enrolled under the 'subject enrolment only' status[7], we were interested in how these two categories related. Initially, we graphed both categories to visualise the relationship (figure 4.15).

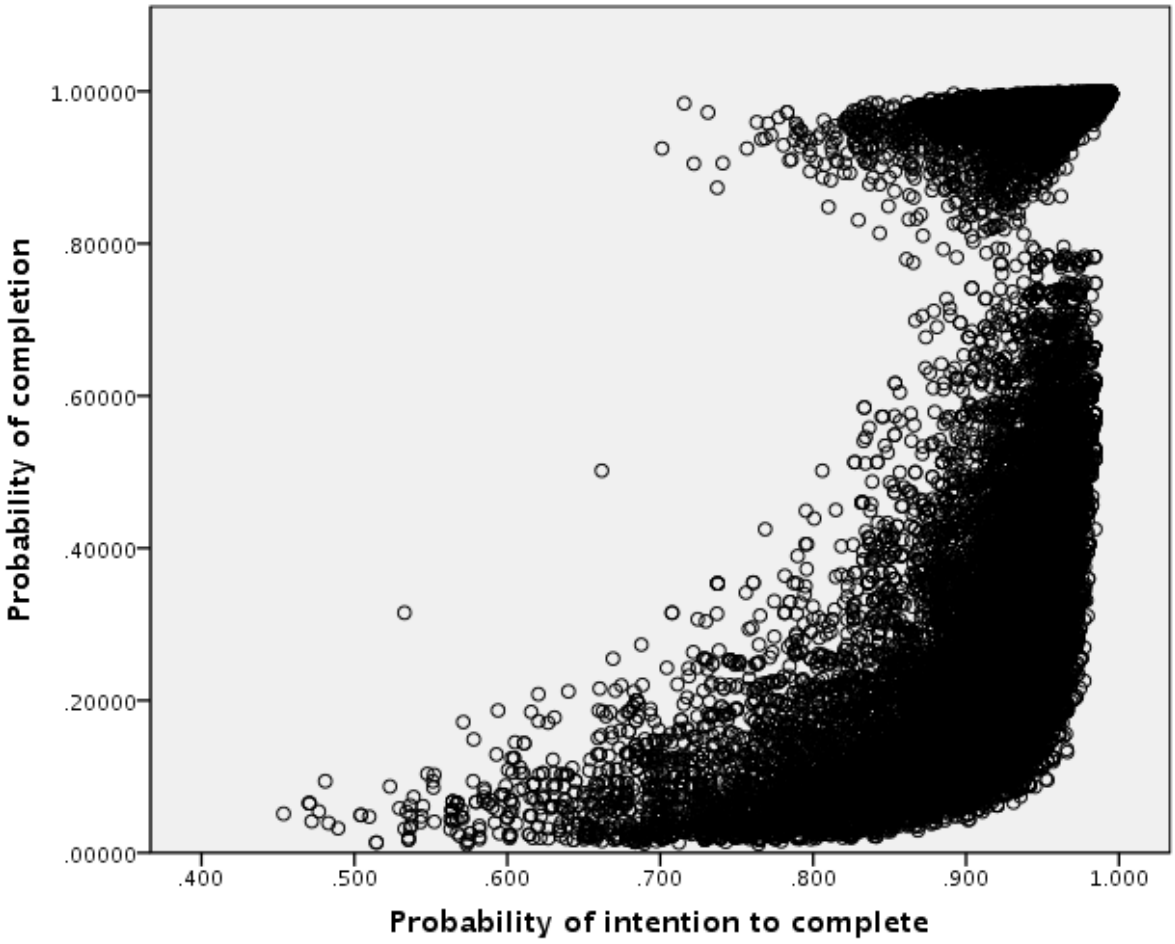

Figure 4.15 Probability of completing v probability of intention to complete

The relationship between completion and the intention to complete was not linear. There appeared to be two subgroups, separated at a probability to complete of about 0.8. The first subgroup, located between a probability of completion of zero and about 0.8, revealed the pattern of a rational function, where, as the probability of intention to complete increases, the probability of

[7] The observations were removed as completion is not possible for these students

completion increases slowly to a certain point and then rapidly over a small increase in the probability of intention to complete. The second subgroup with a probability to complete of above 0.8 exhibited a distinctly different pattern. Here we found a funnel-like shape where probability of intending to complete and probability to complete increase at similar rates, and culminate close to the 1-1 nexus of the graph. Further analysis revealed that the first subgroup is representative of part-time students and the second subgroup of full-time students. In the part-time subgroup there were large numbers of students with a low probability to complete and a much smaller proportion of students with a high probability to complete, while most still had a fairly strong intention to complete. Among full-time students the probability of intending to complete and probability of actual completion was universally strong. Overall the high propensity of intending to complete was no surprise, as in the 2011 SIS 93% of all students stated their intention to complete. Values for the probability of completing on the other hand were distributed across almost the entire spectrum between 0 and 100%, including a substantial number of students exhibiting close to zero probability to complete.

The unique pattern of the probability of completing and the probability of intention to complete necessitated that the subgroups of full-time and part-time students were examined in separate analyses.

Part-time student completions

In the first part of this paper we investigated the benefit to completion. It was thus a logical extension of the theme to examine the impact of not only the intention to complete, but also the benefit of completion on actual completions. Initially, we had a look at the correlations between the pay-offs, the probability of intention to complete and the probability of completing (table 4.18). Due to the non-linear relationship between some of the items we used a non-parametric technique, Spearman's rank order method, to

determine correlation coefficients. For this analysis and all following analyses in this section we used the subset of part-time students.

Table 4.18 Correlations between payoffs and completion probabilities (part-time students)

Spearman correlations – part time students

		Employment	Salary	Occupational level	Satisfaction	Further study	Improved employment conditions	Probability of completing	Probability of intention to complete
		Pay-off							
Employment	**Pay-off**	1							
Salary		.31**	1						
Occupational level		.33**	.78**	1					
Satisfaction		.40**	.30**	.16**	1				
Further study		.09**	-.26**	-.17**	.06**	1			
Improved empl. cond.		.70**	.21**	.23**	.31**	.19**	1		
Probability of completing		.23**	.47**	.39**	.54**	-.30**	.36**	1	
Probability of intention to complete		.30**	.37**	.38**	.27**	-.12**	.37**	.74**	1

** indicates p<0.01

All pay-off and completion scenarios were significantly correlated with one another, with the strongest correlations between the employment pay-off and the improved employment condition pay-off (0.7), the satisfaction pay-off and the employment pay-off (0.4), satisfaction and probability of completing (0.54), salary pay-off and the probability of completing (0.47) and the salary pay-off and the occupational level pay-off (0.78). The correlation between the probability of intention to complete and the probability of completing was 0.74.

We also performed a multiple OLS regression on the probability of completion to account for the impact of individual predictors when the remaining predictors are held constant. The relationship of these two variables for the subgroup of part-time students can be seen in figure 4.16.

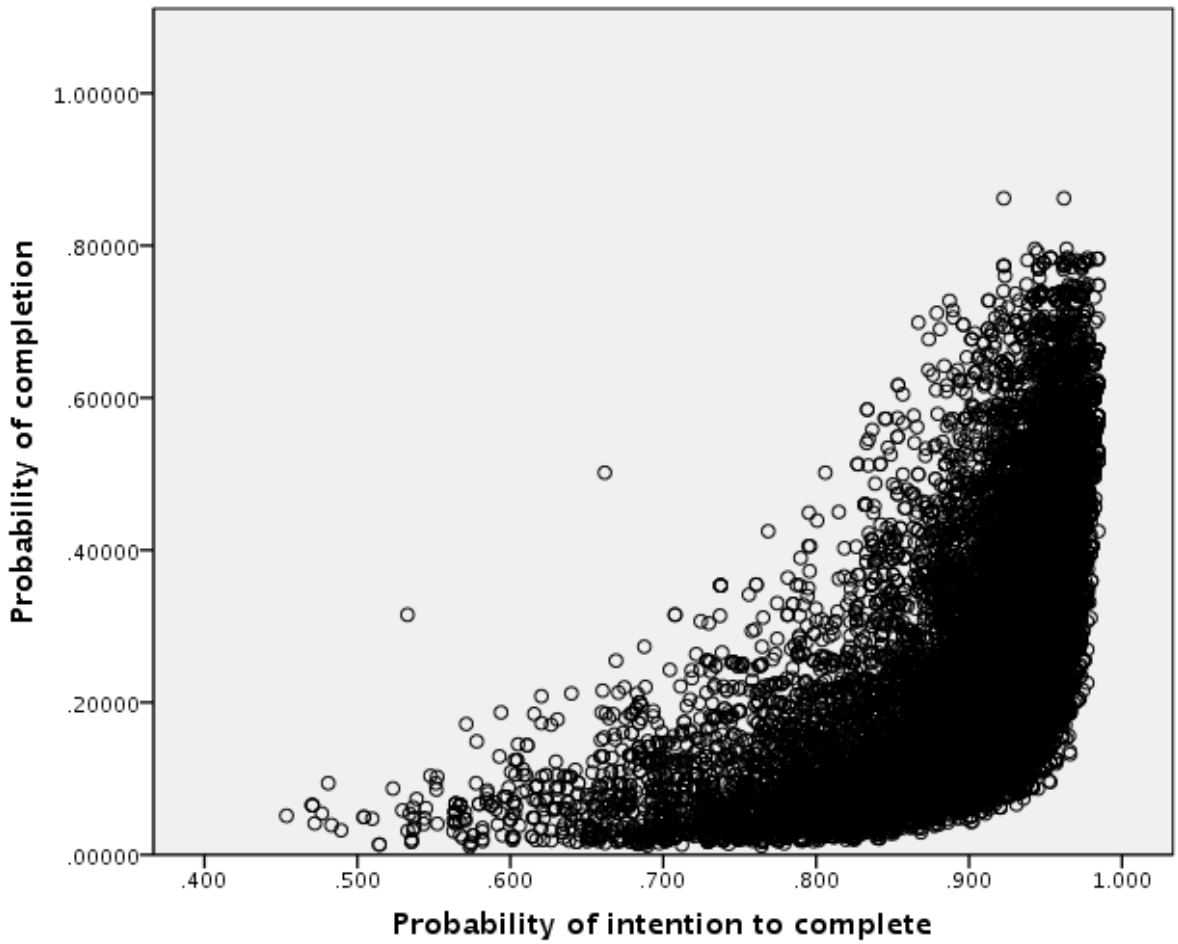

Figure 4.16 Probability of completing v probability of intention to complete, part-time students only

Due to the non-linear relationship between the probability of intention to complete and the probability of completing we decided to transform the variable indicating the probability of intention to complete. Visual inspection of figure 4.16 revealed the pattern of a rational function, characterised by this functional form (3):

$$T = \frac{C_1}{(1-Prob_{completion-intent})} \qquad (3)$$

In this formula T represented the transformed variable and C_1 the parameter that had to be determined via a non-linear regression, using sequential quadratic programming algorithm. C_1 was estimated as 0.016 by the algorithm. Entering this value into the transformation formula and regressing the transformed probability of intention to complete variable on the probability to

complete variable improved the fit of the regression estimation from an r-squared of 0.36 (linear) to 0.47 (transformed).

We used the SOS once again and aimed to employ the probability of completing as the dependent variable and the transformed probability of intention to complete variable as well as the benefits of completion estimated earlier as predictors in an OLS regression. While it appeared sensible to speculate that intention to complete and the pay-offs related to completion played a major role in modelling completion, we also expected that students' direct experiences with their training could contribute to the probability of completion. Previous research in post-compulsory education has shown that satisfaction with the training was an important factor in non-completion (Yorke ,2000; Martinez, 2001). To determine the impact of the probability of intention to complete and the payoffs to completion on the probability to complete and also to evaluate whether student satisfaction with the training can provide additional power to improve the explanatory value of this model we performed a hierarchical regression. In this model we used two sets of predictors: The first set comprised the probability of intention to complete and the benefits to completion (the 'Reduced model'), and second set added the three categories of student satisfaction from the student outcome survey, for example, satisfaction with teaching, assessment, and general learning (the 'Full model'). The resulting model statistics can be seen in table 4.19.

Table 4.19 Impact of intention to complete and pay-offs on completion (part-time students)

			Coefficient	Std error	Beta	t	Sig
Reduced model	Prob. intention to complete		0.53	<0.01	0.51	358.8	<0.01
	Employment	Payoff	-0.19	<0.01	-0.25	-128.6	<0.01
	FT salary		0.49	<0.01	0.29	145.9	<0.01
	Occupational status		-0.13	<0.01	-0.08	-42.1	<0.01
	Further study		-0.05	<0.01	-0.19	-146.3	<0.01
	Improved empl. condition		0.32	<0.01	0.34	163.6	<0.01
	Constant		*-0.31*	*<0.01*		*-116.2*	*<0.01*
				R-squared (adj) = 0.59	F = 67,153		
Full model	Prob. intention to complete		0.52	<0.01	0.50	356.9	<0.01
	Employment	Payoff	-0.19	<0.01	-0.26	-132.9	<0.01
	FT salary		0.50	<0.01	0.29	147.1	<0.01
	Occupational status		-0.14	<0.01	-0.09	-44.6	<0.01
	Further study		-0.05	<0.01	-0.19	-145.1	<0.01
	Improved empl. condition		0.32	<0.01	0.34	165.7	<0.01
	Teaching	Satis.	-0.02	<0.01	-0.09	-47.4	<0.01
	Assessment		0.01	<0.01	-0.01	-3.2	<0.01
	General		0.02	<0.01	0.09	54.7	<0.01
	Constant		*-0.27*	*<0.01*		*-92.4*	*<0.01*
				R-squared (adj) = 0.6	F = 46,017		

The reduced model explained the probability of completing very well with the predictors explaining about 59% of the variance in the probability of completing. All predictors were significant, although standardised regression coefficients suggested that the strongest impact on completion by far came from the probability of intention to complete. Surprisingly, the employment, occupational status and further study pay-offs had a negative, albeit in the case of occupational status fairly limited impact on the probability of completion, suggesting that these three pay-off considerations played little substantive role when students decided not to complete their qualification. On the other hand, pay-offs from salary and improved employment conditions were positively and strongly associated with completion.

The full model added the satisfaction component to the reduced model. As the change in R-squared between the reduced and full model was rather small (0.59 v 0.6), a partial F test (4) was computed to determine whether the

addition of the satisfaction component made a significant difference to the model. We calculated the partial F-test via:

$$F = \frac{(SSQ_{res_{red}} SSQ_{res_{full}})/3}{MSS_{res_{full}}} \quad (4)$$

which yielded and F = 1,522 ($p<0.01$), indicating that the addition of the satisfaction component indeed made a significant contribution to the model. General satisfaction had the strongest positive impact of the three satisfaction categories, with higher general satisfaction related to stronger probability of completion. Teaching satisfaction was negatively associated with completion and the impact of satisfaction with assessment was negligible. While the association of increasing teaching satisfaction with a higher completion probability may appear implausible at first, Edwards and Waters (1982) have also found a reduced role of direct course-related satisfaction with course completion in post-secondary education. In their analysis it was also general satisfaction that exhibited much stronger predictive power for completion.

We can conclude from this section that the intention to complete early in the training provides the strongest indicator of actual completion for part-time students. The impact of individual completion payoffs on completion is mixed. While the payoff in terms of full-time salary and improved employment conditions appears positively related to the probability of completion other payoffs investigated play no positive role. Of the satisfaction items it was mainly general satisfaction that was associated with positive completion outcomes for part-time students.

Completion deficits of part time students

We already pointed out earlier that there was a striking discrepancy between the relatively strong overall intention to complete a qualification and dismal actual completion. Some commentators have argued that, in addition to ‘generic’ reasons for non-completion (for example, personal circumstances,

financial difficulties, sickness, dissatisfaction Yorke 1997, 1999) there are also VET-specific reasons. The main VET-specific non-completion reason cited in this context was that a substantive number of students only aimed to gain specific skills rather than a full qualification and thus discontinued their studies once those skills had been acquired (Karmel & Nguyen, 2006; Snell & Hart, 2007). While this contention may explain the low actual completion rates, it was still not sufficient to explain the comparatively large proportion of students who appeared to commence their studies with the intention to complete.

To shine a light on this disparity we developed the concept of completion deficit (CD, (5)). We defined the completion deficit as the difference between the probability of completing and the probability of intention to complete.

$$CD = P[Compl_{intent}]-P[Compl] \quad (5)$$

The range of possible values for this new variable was (-1, 1) where -1 indicated someone who did not intend to complete but did, and 1 denotes a student who intended to complete but did not. Given the strong pattern of part-time students who exhibited a strong probability of intention to complete coupled with a low probability of actual completion (see figure 4.16), we expected the mean of the completion deficit for this student type to be well above 0. The distribution of the CD for part-time students can be seen in figure 4.17.

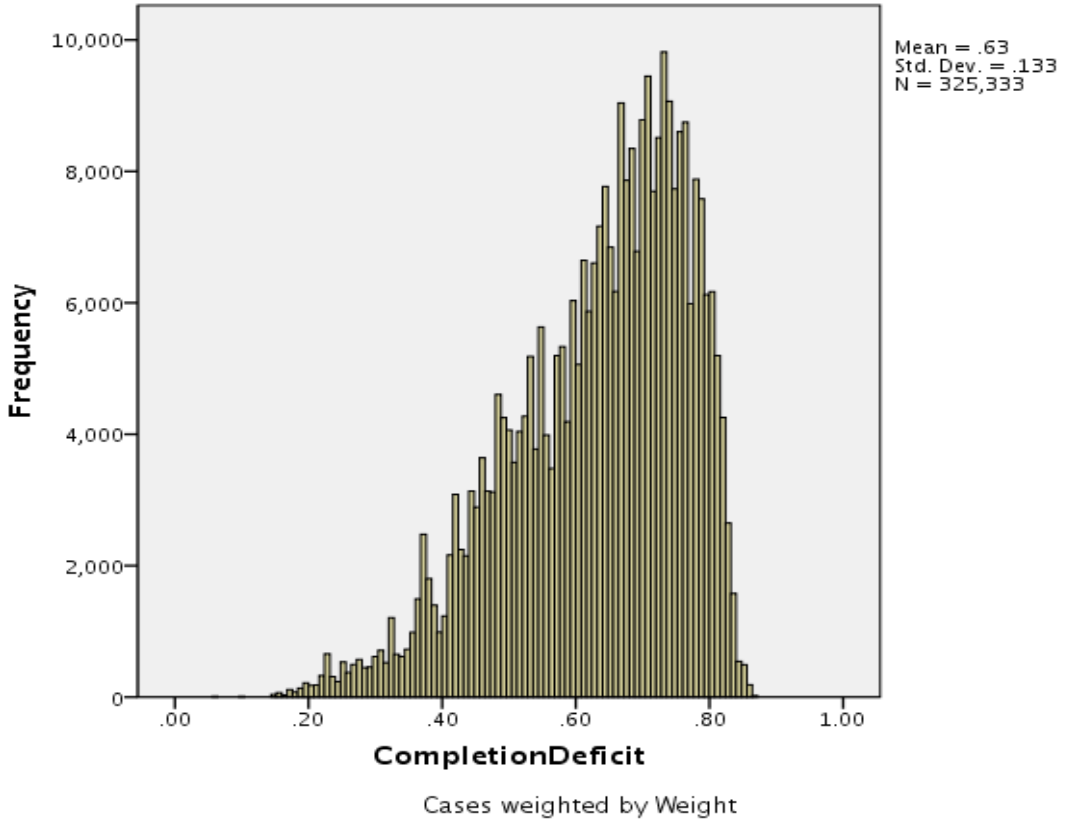

Figure 4.17 Histogram, completion deficit, part-time students

In the completion deficit graph we found no students with a negative completion deficit. On the right from zero, starting at about 0.15 and with increasing student numbers as the CD increases, were those students who exhibited a positive completion deficit, for example, those who were more likely to intend to complete than to actually complete. The number of such students continues to increase up to a completion deficit of about 0.7 and then drops rapidly. The students located between about 0.4 and 0.8 appeared to represent a significant share of those students who account for the low completion rates in VET. From a policy standpoint that favours an increase in completion rates this must be considered significant because of the magnitude of the completion deficit and the large number of students exhibiting such substantial completion deficits. From an administrative viewpoint, these students who exhibit such large completion deficits could also represent those who should potentially be targeted in any intervention measures with the aim to increase completion rates. A conceivable benefit of utilisation of the

methodology outlined above would be that a completion deficit score could be computed for individual students at the point of enrolment.

The advantage of the completion deficit over the probability to complete in analysing is that the completion deficit puts emphasis on those who have a high probability of intention to complete but ultimately a low probability of actual completion. We performed a hierarchical OLS regression to get a basic idea of how the variables we have developed in this paper interrelate with the completion deficit. We used the pay-offs for employment, salary, occupational status, further study and improved employment conditions as well as the different types of student satisfaction as predictors and the completion deficit as dependent variable. Again our secondary aim in this analysis was to determine whether student satisfaction with the training played an additional role in completion (table 4.20).

Table 4.20 Impact of payoffs and satisfaction on completion deficit (part-time students)

			Coefficient	Std error	Beta	t	Sig
Reduced model	Employment	**Payoff**	0.20	<0.01	0.33	130.2	<0.01
	FT salary		-0.56	<0.01	-0.40	-151.9	<0.01
	Occupational status		0.15	<0.01	0.12	46.1	<0.01
	Further study		0.05	<0.01	0.23	131.7	<0.01
	Improved empl. condition		-0.37	<0.01	-0.47	-184.2	<0.01
	Constant		*1.17*	*<0.01*		*423.6*	*<0.01*
				R-squared (adj) = 0.289	F = 22,493		
Full model	Employment	**Payoff**	0.21	<0.01	0.33	134.3	<0.01
	FT salary		-0.56	<0.01	-0.40	-153.6	<0.01
	Occupational status		0.16	<0.01	0.13	50.0	<0.01
	Further study		0.05	<0.01	0.22	130.0	<0.01
	Improved empl. condition		-0.36	<0.01	-0.46	-185.2	<0.01
	Teaching	**Satis.**	0.02	<0.01	0.13	49.7	<0.01
	Assessment		<0.01	<0.01	0.02	6.2	0.77
	General		-0.02	<0.01	-0.14	-64.7	<0.01
	Constant		*1.12*	*<0.01*		*367.0*	*<0.01*
				R-squared (adj) = 0.304	F = 15,071		

The reduced model explained about 29 % of the variance in the completion deficit. Completion payoffs from salary and improved employment conditions were most strongly associated with the completion deficit. In both cases a

lower payoff indicated a higher completion deficit, suggesting that a low payoff for these two variables is related to a higher probability of intention to complete paired with a lower probability of actual completion. The impact of payoffs for employment, occupational status and further status on the completion deficit was of smaller magnitude and positive. We also wanted to know whether the addition of the satisfaction component improved the completion deficit model significantly for part-time students. The resulting full model had a slightly higher r-squared of around 0.30, indicating the model improved only moderately. However, the partial F test yielded a result of 1,918 ($p<0.01$), indicating that the added satisfaction component improved the model significantly. General and teaching satisfaction had opposite effects on the completion deficit with increasing teaching satisfaction related to a higher completion deficit, and thus a bigger rift between the intention to complete and actual completion. On the other hand, higher general satisfaction indicated a reduced completion deficit. Our results thus suggest that an improved overall general VET experience for students may be more helpful to reduce the completion deficit than an increase in teaching satisfaction.

Finally, we wanted to identify those groups of part-time students who exhibited particularly striking completion deficits in order to enable the easy identification of 'high risk' groups based on their demographic and educational characteristics. Chi-squared Automated Interaction Detection was once again employed to calculate and visualise the classification of groups of part-time students. Independent variables in this analysis were the same as in the earlier investigation on completion pay-offs: age group, gender, qualification level, prior education, employment status before training, remoteness, socioeconomic status, field of education and study status. The resulting classification tree can be seen in figure 4.18.

Starting with an overall completion deficit of 0.63 for this student type, the first important split is at the qualification level: Here, it can once again be seen that Certificate I/II has the most poorest outcome with a completion deficit of 0.72, whereas Certificates III/IV and Diploma and above qualifications completion deficits were estimated as 0.62 and 0.46 respectively. While node splits after the qualification levels were not uniform across qualifications several patterns emerged. Of those, most notable were that the completion deficit increases with age, males have higher completion deficits than females, and rural and regional students have higher completion deficits than urban students. We can thus generalise that at the completion deficit extremes of part-time students there are young female urban diploma and above students likely to display relatively low completion deficit, whereas older, male, Certificate I/II students from remote areas are more likely to exhibit high completion deficits.

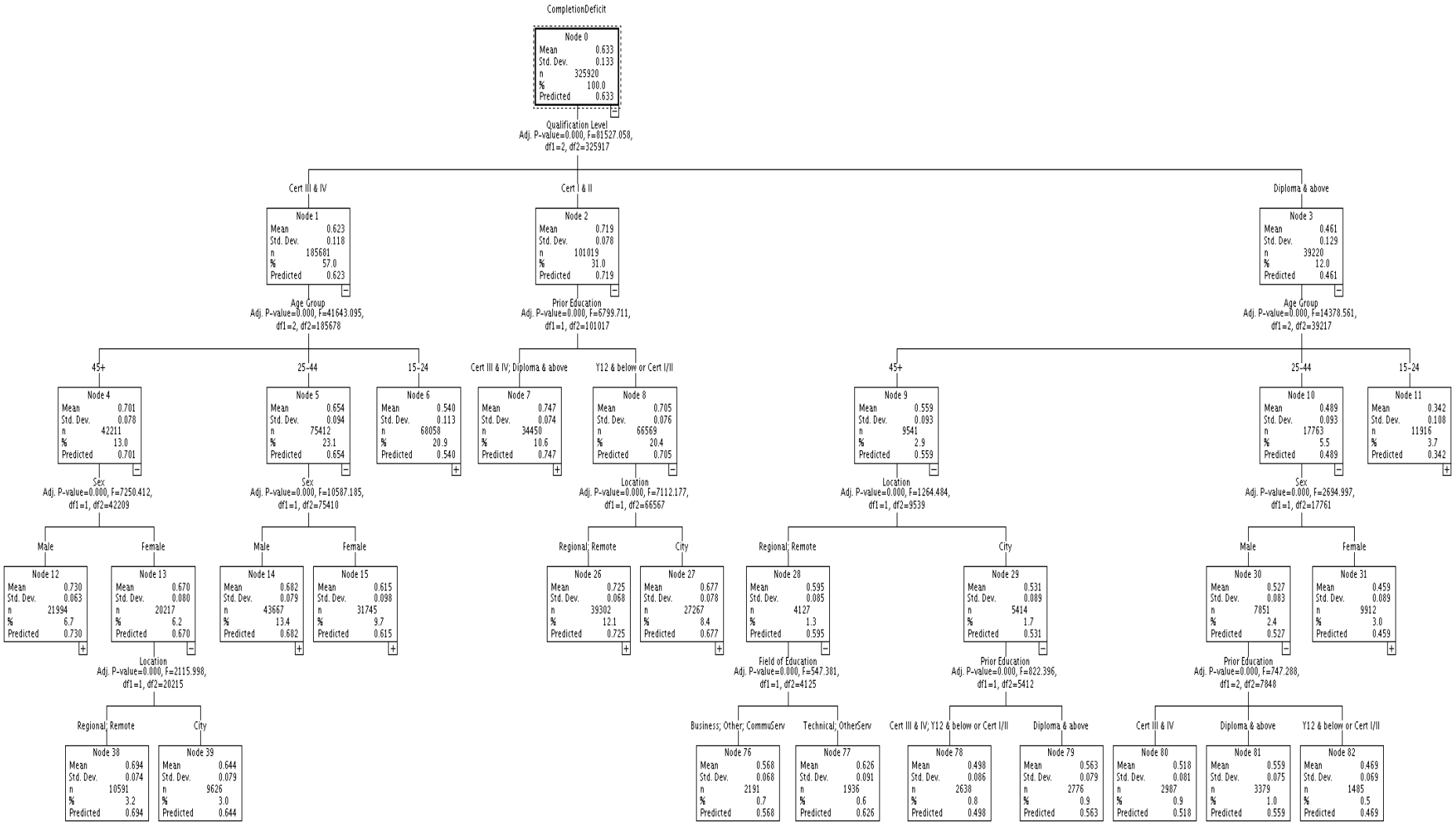

Figure 4.18 CHAID diagram, completion deficit, part-time students

Full-time student completions

Full-time students' patterns of the probability of completion and the probability of intention to complete were substantially different from part-time students (see figure 4.15). Both probabilities were on average substantially higher, and they appeared to be more linearly related. As the shape and location of the full-time student population in respect to completion and intention of completion was substantially different from the part-time student population, we were interested whether there would also be differences between these two groups of students in terms of correlations between the main variables analysed in the study. Correlations are displayed in table 4.21.

Table 4.21 Correlations between payoffs and completion probabilities (full-time students)

Spearman correlations – full-time students		Employment	Salary	Occupational level	Satisfaction	Further study	Improved employment conditions	Probability of completing	Probability of intention to complete
		Pay-off							
Employment	**Pay-off**	1							
Salary		.12**	1						
Occupational level		.20**	.72**	1					
Satisfaction		.33**	.22**	.08**	1				
Further study		.26**	-.15**	-.08**	.13**	1			
Improved empl. cond.		.52**	.19**	.27**	.17**	.06**	1		
Probability of completing		.01*	.42**	.32**	.43**	-.21**	.40**	1	
Probability of intention to complete		.21**	.24**	.29**	.11**	.04**	.44**	.62**	1

** indicates p<0.01; * indicates p<0.05

All correlations between payoff variables and completion probabilities for full-time students were significant and also in the same direction as correlations for part-time students, although the correlation between the payoff for employment and the probability of completion was marginal in the case of

full-time students. The most noticeable difference between both student types was that the correlations between payoffs and the probability of intention to complete and probability of completion were generally of lower magnitude in the part-time student subgroup, with the exception of the correlation between the payoff for improved employment conditions and the probabilities of completion and intention to complete.

We again performed a multiple OLS regression on the probability of completion to account for the impact of individual predictors when the remaining predictors are held constant. Given the unique relationship between the probability of intention to complete and the probability of completion (recall figure 4.15), we took a closer look at this relationship for the subset of full-time students (figure 4.19).

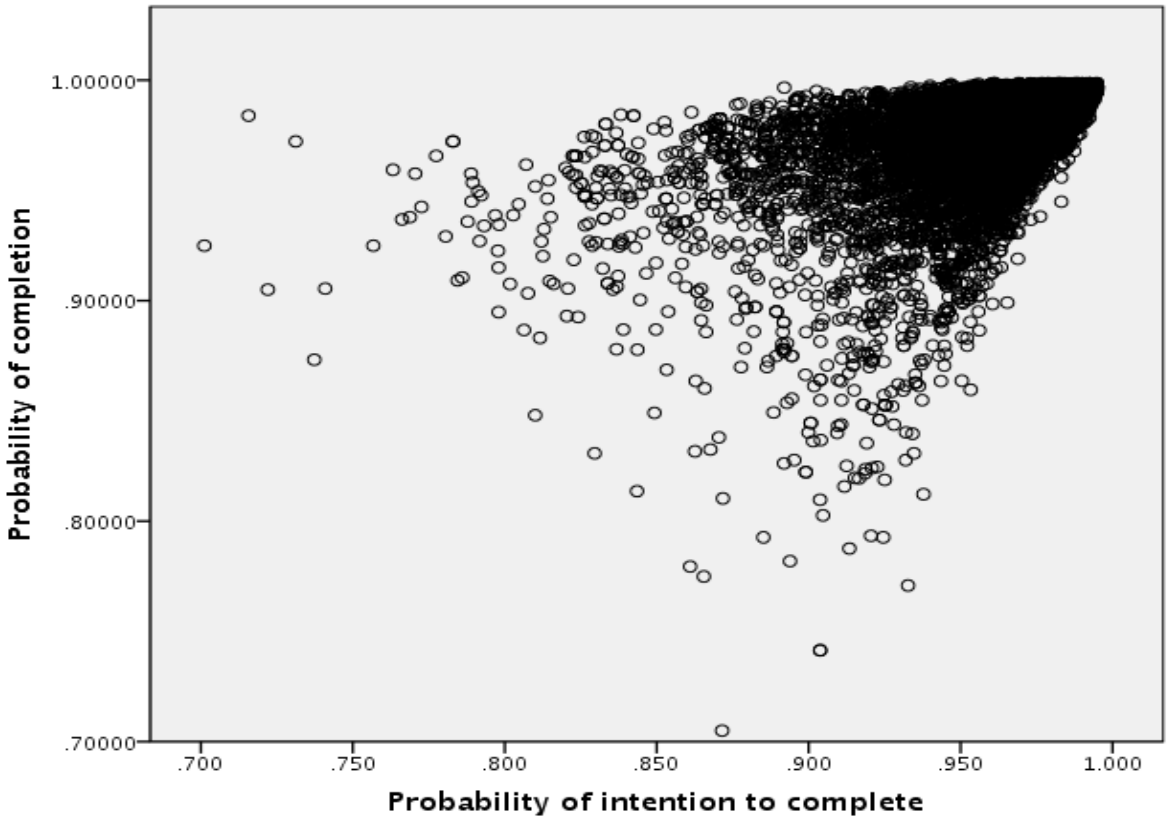

Figure 4.19 Probability of completing v probability of intention to complete, full-time students only

The funnel-like shape toward the [1,1] nexus in the above graph may foreshadow potential heteroskedasticity problems when subsequently regressing the probability of intention to complete on the probability of completion. We performed such a regression and created a plot of the standardized residuals versus the standardised predicted values (figure 4.20).

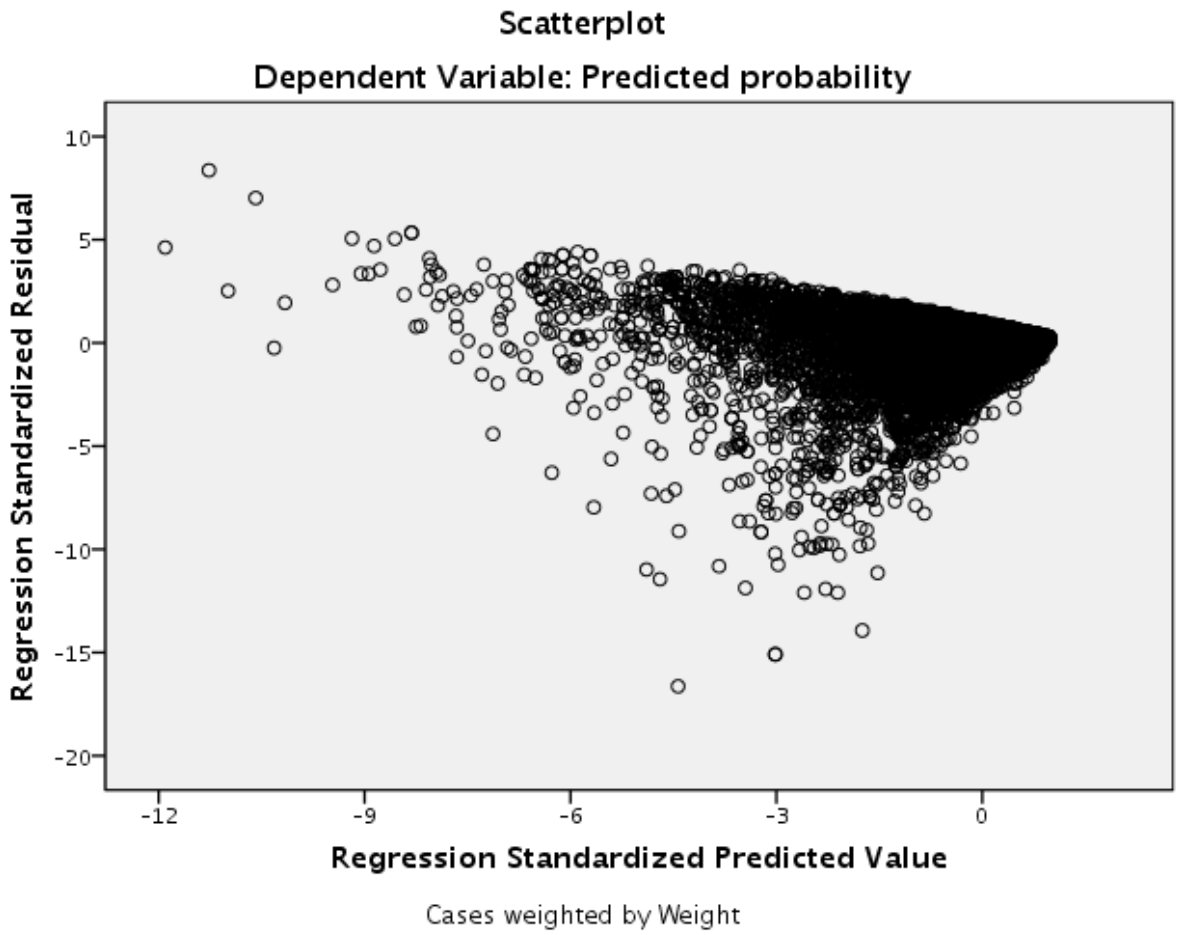

Figure 4.20 Standardised residuals v standardised predicted values

Visual inspection of this plot confirmed the heteroskedasticity suspicion. While heteroskedasticity does not bias estimated regression coefficients, it does have an impact on standard errors. Therefore we used the 'robust' standard error option in subsequent regressions where the probability of intention to complete was used as a regressor on the probability to complete.

Table 4.22 Impact of intention to complete and pay-offs on completion (full-time students)

			Coefficient	Std error	Beta	t	Sig
Reduced model	Prob. intention to complete		0.37	0.01	0.48	259.8	<0.01
	Employment	Payoff	-0.01	<0.01	-0.12	-55.6	<0.01
	FT salary		0.05	<0.01	0.25	104.5	<0.01
	Occupational status		-0.02	<0.01	-0.13	-52.9	<0.01
	Further study		-0.01	<0.01	-0.16	-92.6	<0.01
	Improve empl. condition		0.03	<0.01	0.22	99.0	<0.01
	Constant		*0.59*	*0.01*		*473.9*	*<0.01*
				R-squared (adj) = 0.456	F = 29,614		
Full model	Prob. intention to complete		0.37	0.01	0.48	260.5	<0.01
	Employment	Payoff	-0.01	<0.01	-0.12	-56.2	<0.01
	FT salary		0.05	<0.01	0.25	104.8	<0.01
	Occupational status		-0.02	<0.01	-0.13	-53.8	<0.01
	Further study		0.00	<0.01	-0.16	-91.9	<0.01
	Improved empl. condition		0.03	<0.01	0.22	98.2	<0.01
	Teaching	Satis.	-0.01	<0.01	-0.04	-15.8	<0.01
	Assessment		0.00	<0.01	0.00	0.3	0.77
	General		0.01	<0.01	0.05	23.8	<0.01
	Constant		*0.59*	*0.01*		*468.4*	*<0.01*
				R-squared (adj) = 0.458	F = 19,880		

The reduced model predicting the probability of completion had a slightly less explanatory value for the subset of full-time students (r-squared = 0.46) than the identical model for part-time students (r-squared 0.59). The probability of intention to complete was again the strongest predictor while positive completion payoffs from salary and improved employment conditions are also the strongest completion payoff predictors. Completion payoffs from employment, occupational status and further study played a less pronounced role and did not relate to a higher probability of completion. Adding the satisfaction component to the reduced model increased the explained variance only marginally (r-squared from 0.456 to 0.458). However, the partial F-test (F = 225; p<0.01) indicated that the satisfaction component added significantly to the model. As in the model of part-time students, only general satisfaction was positively related to a higher probability of completion, albeit only to a very small degree.

Completion deficits of full-time students

Given the shape of the probability of intention to complete and probability of completion displayed in figure 4.15 we could expect the completion deficit of full-time students to be radically different from the part-time students evaluated in the previous section. The histogram in figure 4.21 shows that this suspicion was warranted.

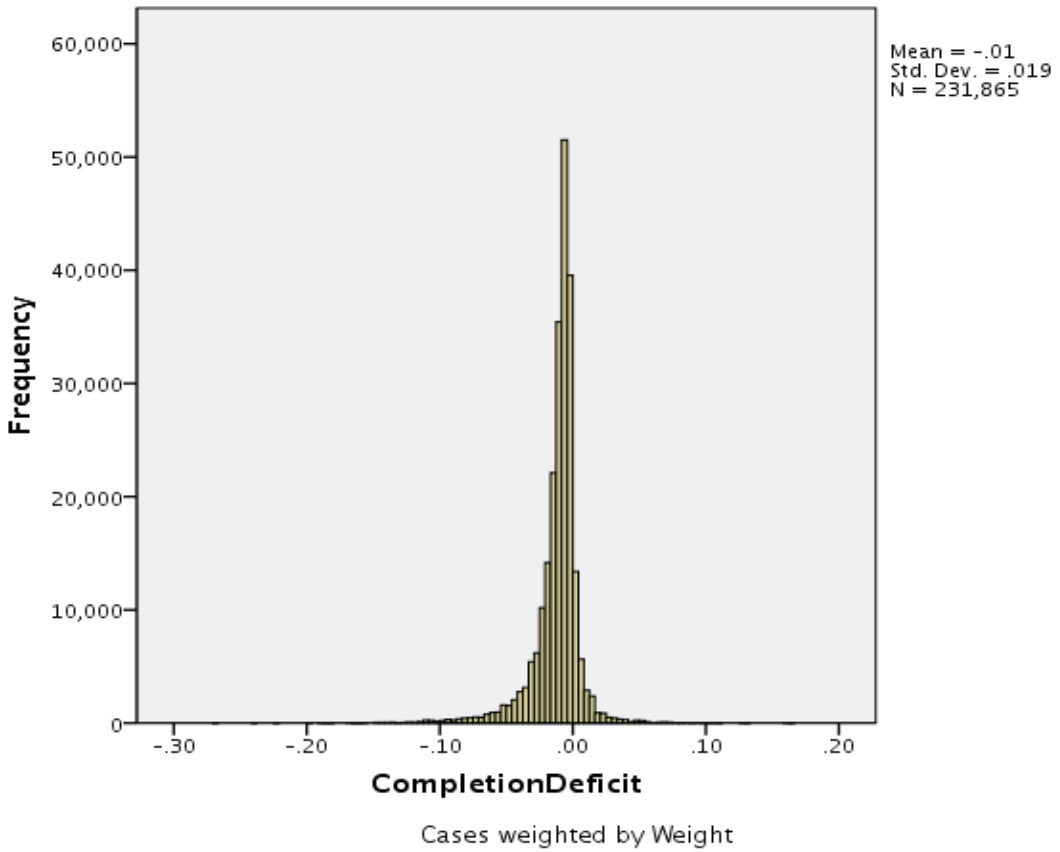

Figure 4.21 Histogram completion deficit full-time students

The histogram displayed a near normal shaped distribution emerging from about -0.1 to 0.05, and centred slightly to the left of zero. This represented an important discovery as it suggests that the problem with substantial completion deficits lies almost entirely with part-time students. Students with a CD value below zero obviously did not have a completion deficit as such as they, against their own expectations, were likely to complete their training despite not intending to do so. On the right hand side of zero was a small

minority of students whose likelihood to intend to complete slightly outweighed the likelihood to actually complete. It would thus appear that generally full-time students do not warrant significant attention from VET administrators aiming to reduce completion deficits. It still remained of interest which determinants predicted differences in the completion deficit for full-time students. We thus repeated our earlier model and regressed completion payoffs and student satisfaction on the completion deficit of full-time students (table 4.23).

Table 4.23 Impact of payoffs and satisfaction on completion deficit (full-time students)

			Coefficient	Std error	Beta	t	Sig
Reduced model	Employment	**Payoff**	-0.01	<0.01	-0.12	-44.0	<0.01
	FT salary		-0.04	<0.01	-0.20	-66.1	<0.01
	Occupational status		0.05	<0.01	0.28	90.0	<0.01
	Further study		0.01	<0.01	0.18	81.9	<0.01
	Improved empl. condition		0.02	<0.01	0.18	66.5	<0.01
	Constant		*-0.05*	*<0.01*		*-101.7*	*<0.01*
				R-squared (adj) = 0.086	F = 3,993		
Full model	Employment	**Payoff**	-0.01	<0.01	-0.11	-43.0	<0.01
	FT salary		-0.04	<0.01	-0.21	-66.8	<0.01
	Occupational status		0.05	<0.01	0.28	91.4	<0.01
	Further study		0.01	<0.01	0.18	81.2	<0.01
	Improved empl. condition		0.02	<0.01	0.18	67.4	<0.01
	Teaching	**Satis.**	0.00	<0.01	0.02	6.5	<0.01
	Assessment		0.00	<0.01	0.03	10.4	<0.01
	General		<-0.01	<0.01	-0.07	-26.8	<0.01
	Constant		*-0.05*	*<0.01*		*-91.1*	*<0.01*
				R-squared (adj) = 0.089	F = 2,595		

The explanatory power of this model was modest (r-squared reduced model = 0.086; full model = 0.089). The strongest predictor was the completion payoff from occupational status and indicated that higher payoffs are related to increased completion deficits. The satisfaction component improved the model significantly (partial F-test F = 241, p<0.01), albeit to a small degree. Decreasing general satisfaction was associated with higher completion deficits.

To determine which educational and demographic variables were related to completion deficits of full-time students we created a CHAID diagram with completion deficit as the outcome variable (figure 4.22). Due to the narrow distribution of the completion deficit the differences between nodes in the CHAID tree were minuscule. We thus collapsed most categories and retained only those branches that had recognisable differences. The first level of significant predictors was the qualification level, where Certificates III and IV had the highest completion deficit (-0.008) albeit only marginally different from Certificate I and II (-0.011) and Diploma or higher qualifications (-0.024). Differences became even more minuscule at the following levels, and only some distinct sub-groups produced an actual positive completion deficit (for instance 'Other services' field of education at the Certificate I and II level, or 45+ aged students, in the 'Technical' and 'Community services' field of education at the Certificate I and II level). While it was clear that there was a very low distinguishability present between the various CHAID categories it needs to be kept in mind that in practical terms the absence of a general completion deficit problem among the full time student population makes the identification of subgroups a purely theoretical problem.

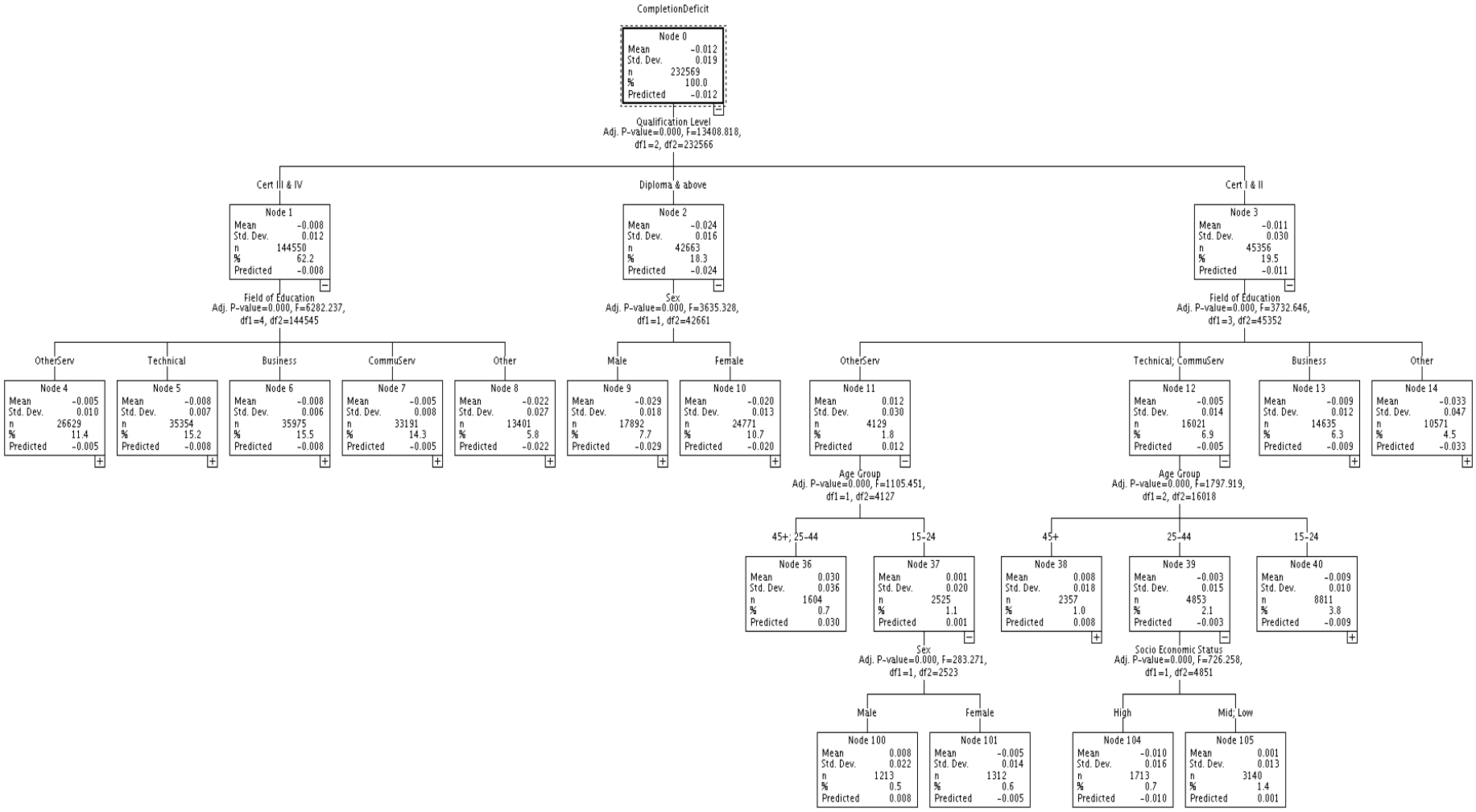

Figure 4.22 CHAID diagram, completion deficit, full-time students

4.1.4 Conclusion

In this analysis we have been able to quantify the benefit of completion for individual students, and aggregated these benefits to student groups of interest. This has yielded some new insights and some results that have confirmed earlier research. As in any research of this type, there were some qualifications that need to be pointed out in order to add an element of caution when interpreting these results. While certain standard caveats apply that point to shortcomings inherent to research involving surveys (such as non-response bias, sampling error, non-sampling error etc.), the main issue that should be kept in mind when pondering the results presented here is that we essentially made a distinction between module completers and graduates of VET training programs and classified them as either non-completers and completers. While this was essentially true, we omitted those students from our analysis who discontinued their studies before completing their first module, predominantly because we had virtually no data on them. Given that module completion rates averaged around 90%, we can assume that this group of non-completers who had not completed a single module was very small and thus would not have exerted a substantial impact on the results of this study had they been included. It also needs to be re-emphasized that the respondent numbers for the salary outcome were only about a third of those who responded to the other questions. Another limitation in the analysis of the satisfaction outcome was that the predictors only accounted for a very small amount of the variance in the satisfaction outcome, indicating that there were unknown factors that we did not observe in the SOS which impacted on student satisfaction. We were able to identify that the largest payoffs from completion of a qualification occur for the outcomes 'Further study after training', 'Improved employment conditions', and 'Employment after training'. For those outcomes we could also identify which student groups

benefited most and least in respect to these payoffs. The respective benefits from completion at times differed substantially between groups of students.

In the second part of this paper we analysed the gulf between intended and actual completions in the VET sector. As there were no data available that tracked a cohort of students from enrolment to completion with respect to their intentions, we had to develop models that predicted the probability of intending to complete for those students where we knew the completion outcome from the Student Outcome Survey. It needs to be kept in mind that our analysis was based on such modelled data rather than actual data.

In this paper we examined issues of completion of qualifications in the VET sector. This included the benefit of completion, the impact of the benefit to completion on completions and the remarkable rift between intended and actual completions. We have been able to identify training outcomes where the pay-off to completion was substantial (for example, for further study) and where completion mattered little (for example, salaries, occupational status and satisfaction). Furthermore, our chosen methodology enabled us to identify educational and demographic characteristics of groups of students for whom completion mattered most and least. Subsequently we investigated the VET sector's low completion rates and the stark contrast that existed in respect to students' intentions to complete. The main finding here was that completion issues in the VET sector are predominantly a problem within the part-time student population. Our analysis also revealed a relationship between completion and the intention to complete, and to a weaker extent the pay-off to completion. Finally, we developed the notion of the completion deficit. This concept was then used to develop a classification of students that exhibited a particularly strong disparity between intended and actual completion. One unexpected outcome of our analysis was the relatively small contribution of student satisfaction to the probability of completion and the

completion deficit. While we could confirm the stronger magnitude of 'general' satisfaction over course-related satisfaction found in earlier research in respect to completion outcomes, it was still striking to observe the relatively low contribution of overall satisfaction to our models dealing with completion. Furthermore, it is possible that, due to the unique properties of the vocational education system discussed earlier in this paper (for instance, that a substantial proportion of non-completers fail to graduate because they have gained the skills they set out to acquire, or non-completion due to arising employment opportunities during the training), satisfaction is indeed less related to completion in the VET system when compared with the higher education sector. Furthermore, it is possible that the small contribution of satisfaction to the probability of completion and the completion deficit may partly reflect deficiencies in the way satisfaction was measured.

This paper's intention was threefold. Firstly, it was intended to examine contemporary and controversial issues in vocational education and training surrounding completion. As such we have been able to report on the current state of affairs in the sector using the latest available data. This included the detection of groups of students that exhibit low propensities to successfully complete their qualifications. It also included the identification of student typologies who may benefit most from completion. Our second aim was to develop and promote quantitative methods that can be used by researchers in the VET sector. In this paper we used a variety of quantitative methods and demonstrated how these methods could be applied to investigate statistical completion patterns in contemporary VET populations. Finally, we wanted to develop practical tools for educational practitioners that can be used with the aim to improve outcomes in vocational education. For instance, applying the methods developed in this paper it would be possible to identify students at high risk of non-completion. These could be either students having low probabilities of completion, or students that have a high completion deficit

score. The main benefit of the methods developed in this paper is that such scores could be calculated and assigned to students at the time of enrolment. It would thus be possible to potentially 'follow' students at risk of non-completion and aim potential interventions at them. One such intervention could be to use the 'benefit to completion', also developed in this paper, as a means of persuasion for wavering students.

5 Efficiency and effectiveness in Australian TAFE institutes – implications of this research: Portfolio paper 4

In the portfolio presented here we studied the performance of the TAFE sector in the provision of government-funded vocational education in Australia. In order to be able to better delineate core aspects of institutional performance we considered the measures of efficiency and effectiveness separately. Major emphasis in this portfolio was on the development and application of methodologies that enabled the creation and application of useful indicators for assessing the performance of individual institutions. This portfolio was divided into three constituent research papers to deal with three separate aspects of TAFE performance. Paper one investigated several conventional 'soft' and 'hard' performance measures of effectiveness at the institutional level, while paying specific attention to, and addressing, some shortcomings in the available data. In keeping with the institutional performance analysis theme, paper two developed two measures of efficiency and investigated issues surrounding the transformation of resources into educational and post-educational outcomes. While returning to the overarching theme of effectiveness, paper three explored the issue of course completion at the system level.

In this linking paper we will briefly review the results of this portfolio and how these results complement each other, consider possible implications of our research, discuss research limitations and take a look at potential pathways for future research.

5.1 What have we learned through this portfolio of research

The first portfolio paper aimed to look at institutional effectiveness, which we defined as a summary performance measure that facilitates the assessment of whether students or stakeholders receive value from their investment in vocational education. We were interested in what type of effectiveness indicators could be developed using data from the Student Outcome Survey and some other auxiliary data sources. Furthermore we developed methodologies that enabled the comparability of institutions disparate in terms of student population and educational profile while also addressing inherent bias in the survey data. The final major aim for this portfolio paper was to explore relationships between the indicators developed as well as the creation of a summary effectiveness measure.

To facilitate the comparability between institutions we made use of statistical techniques that allowed us in our analysis to control for confounders such as demographic composition and educational characteristics of institutional student bodies. We also employed selection models designed to alleviate non-response bias inherent in the Student Outcome Survey. Resulting from our analysis we created institutional effectiveness scores for seven distinct performance categories. The effectiveness indicators developed comprised 'soft' and 'hard' indicators in an effort to balance student-based subjective 'soft' measures (such as satisfaction items) with objectively measurable 'hard' indicators (such as employment outcomes). Specifically, examination of the Student Outcome Survey and publicly available labour market data enabled us to develop seven distinct performance measures for each individual TAFE

institute. We accomplished this via the development of a methodology based on mixed effect models that accounted for the selection effect in the Student Outcome Survey. This methodology also enabled us to identify selection bias in several of the developed indicators. Subsequently, a 'traffic light' evaluation scheme was developed that identified those institutes displaying statistically significant over- or under-performance. In order to better visualise the outcome of this analysis we presented institutional results in caterpillar graphs. To integrate our results from the performance measure investigation we applied cluster analysis to reveal performance profiles against the performance indicators created. Finally, a summary indicator was developed using the Euclidean distance concept. Our overall finding was that there is a large variability in performance within the Australian TAFE system and that institutional under- and over-performance can be identified by using the methodologies presented here in this portfolio. The key advantage of our performance indicators over those raw indicators that are currently used by policymakers is that the indicators developed in this research give a more accurate picture of performance, as they account for differences in institutional, demographic and environmental backgrounds and also address shortcomings of the survey instrument used to create such performance measures.

In our second portfolio paper we sought a meaningful way to characterise efficiency in the context of the TAFE system and set out to analyse efficiency via a parametric and non-parametric method. Efficiency characterises the relationship between inputs and outputs in a process and in the scope of our study we explored how different types of efficiency relate to one another and whether a typology of efficient institutions can be developed. To solve these problems we applied the stochastic frontier method to determine two conceptually different types of institutional efficiency. These efficiencies were teaching efficiency, measuring the efficiency of converting financial and

administrative resources into teaching loads, and employment efficiency, quantifying how well individual institutions transform resources into labour market outcomes of their students. Here, too, we adjusted our models for demographic and institutional confounders. As a result we produced two conceptually different efficiency scores for each institution. While our utilisation of stochastic frontier analysis, a parametric method, has been the first effort to ascertain efficiency in Australian vocational education, we were interested how the outcomes of this method compare to the results of a non-parametric method. For a comparison of results we thus repeated our analysis employing non-parametric data envelopment analysis. This technique had been used before in the assessment of the efficiency in various vocational education settings. The results of our efficiency analysis were used to create a ranking of individual institutions using both competing methodologies. We found that both methods produced similar results but that their application is sensitive to a number of external circumstances. The two different types of efficiencies calculated did not exhibit the properties of a linear relationship. An essential inference from our research is that a significant amount of consideration should go into precisely defining the type of efficiency that needs to be determined. Our study identified a number of explanatory variables for both types of efficiencies, including remoteness of an institute as the strongest predictor for both types of efficiency. From a purely financial perspective, a central research result was that there appears to be a minimum institutional size (2.7 million teaching hours) above which the size of an institute is not an impediment to efficiency. Finally, we developed a classification using the two efficiency indicators which entailed three efficiency patterns: Low teaching hours efficiency/high employment outcome efficiency, high teaching hours efficiency/low employment outcome efficiency, and high teaching hours efficiency and high employment outcome

efficiency. These efficiency patterns were then used in a discriminant analysis to create a typology of efficient institutions.

Our third paper dealt again with effectiveness, albeit at the VET system level and with one particular aspect of effectiveness: Completions in VET. The research questions to examine here were how the individual benefit from completion can be quantified and how any potential benefit is distributed across various predefined groups of students. In addition we set out to explore characteristics of students who fail to complete their chosen course. Completion is a somewhat controversial topic among VET research practitioners due to a variety of reasons, ranging from the difficulty of estimating completion rates, over-resulting discrepancies in estimates, to the varying interpretations thereof. The difficulties in estimating completion rates stem from the absence of a unique student identifier in the VET system. Previous and present attempts to estimate completion rates have thus relied on various assumptions. In our analysis we used the 'graduate' and 'module completer' classifications to define completion outcomes. We then created a methodology to determine the benefit to completion in respect to the study outcomes of employment after training, post-training salary, occupational status after training, satisfaction with the training, enrolment in further studies after training and improved employment conditions after training. Our analysis revealed that, overall, the benefit to completion in respect to salary, occupational status and satisfaction was negligible; however, completion conferred significant benefits in respect to employment, improved employment condition and enrolment in further study. Analysis of demographic and educational subgroups of students uncovered a wide array of variation in respect to how these groups of students draw specific benefits from completion. We then used Student Outcome and Student Intentions survey data to estimate the probabilities to complete and the intention to complete. Our study revealed that the probability to complete was strongly related to

the completion benefit in terms of salary, satisfaction and improved employment conditions. The probability of intending to complete was most strongly related to the benefit from improved employment conditions. Finally, we utilised the probabilities to complete and probabilities of intending to complete to establish the concept of the 'completion deficit'; for example, the difference of these two values as a measure of the divergence between intention and reality to graduate from a VET course. We aggregated the individual 'completion deficit' scores and then created student typologies in relation to various completion deficits. The completion deficit was found to be lowest for students enrolled in diploma or higher courses.

The research results of the three papers presented in this portfolio provide an insight into potential approaches for performance measurement in the Australian VET sector. As such it was attempted to create and evaluate a broad range of performance indicators, ranging from labour market outcomes, student satisfaction, institutional efficiency to completion rates. As it was one of the aims of this portfolio to utilise predominantly publicly available data, the potential breadth of indicators was somewhat constrained by the availability of data of this nature.

While the performance indicators introduced in portfolio papers one and two dealt with a number of conceptually different categories of performance measured at the institutional level, it would be theoretically possible to aggregate institutional data and create an overall measure which could be termed 'vocational education system performance'. This could be useful for some of the indicators (for instance, satisfaction or employment outcomes) where the aggregates could then be compiled in a time series to allow inferences on changes in these indicators over time. Other indicators, such as those dealing with efficiency, would be less suited to aggregation to the

national level as our efficiency measures reveal efficiencies of individual institutes relative to other institutes.

Another avenue of aggregating the performance measures introduced here would be to establish an overall performance index at the institutional level by combining all available indicators into one, effectively assigning a single summary measure to each institution. This could be accomplished by using a Euclidean distance method, similar to the one developed in portfolio paper one. The potential drawback of such an endeavour would be that too many conceptually different indicators may obscure the interpretative value of such a summary indicator. Usefulness of aggregation of all performance indicators created could be further impeded by the multi-dimensionality of our measures, which are lending themselves more to a multidimensional or multivariate profile-based approach. Such profiles could provide more sensitive targets for intervention or potential areas that could serve as 'tie-breakers' for resource allocations to institutions that are equivalent on certain outcome measures. A single performance measure may be ill suited for such purposes.

In this portfolio, we have attempted to view performance from the perspective of educational policymakers or institutional administrators and create indicators that can help them to make better decisions, with a reduced likelihood for conflict and an enhanced likelihood for targeting resources to areas of greatest need or potential for growth. We also strived for our performance measures to enable managers of TAFE institutions to better understand their own place in the TAFE sector and how they may harness certain aspects of their profile for best effect in deploying scarce resources. In order to achieve these aims it appears counterproductive to aggregate all the performance measures developed into one 'super' indicator. In respect to the effectiveness summary indicator it is also advisable to consider it in conjunction with the constituent indicators.

It would be of interest to uncover if and to what extent the institutional indicators created in this portfolio are associated and how they could fit into a general performance measurement framework. We used the institutional summary effectiveness indicator estimated in our first portfolio paper and correlated them with the institutional indicator for teaching and employment efficiency. When reviewing the resulting correlations, one needs to bear in mind that the effectiveness summary indicator is inverted and increasing values indicate decreasing effectiveness, so that a negative correlation of the effectiveness scores with efficiency scores indicates agreement in the direction of both indicators. The resulting correlations can be found in table 5.1.

Table 5.1 Correlations between efficiency and effectiveness indicators

		Effectiveness	**Teaching**	**Employment**
		Score	Efficiency	Efficiency
	n	*65*	*56*	*56*
Effectiveness	*r*	1.00	0.28	-0.42
Score	*p*		0.04	<0.001
Teaching hours	*r*		1.00	-0.11
Efficiency	*p*			0.40
Employment outcome	*r*			1.00
Efficiency	*p*			

From table 5.1 it is apparent that there is a significant positive relationship between institutional effectiveness and institutional ‘employment outcome’ efficiency. It can thus be inferred that effective institutions also tend to be efficient in the conversion of institutional resources into employment outcomes for their graduates and module completers. On the other hand, the correlation between effectiveness and ‘teaching hours’ efficiency is only marginally significant and negative. It is thus difficult to argue that greater efficiency in the provision of teaching load necessarily relates to higher effectiveness. Institutional efficiency in the provision of teaching was not significantly correlated with the efficiency in producing employment outcomes. From this

we could take away that increased funding per teaching hour is not necessarily associated with an increase in employment outcome efficiency.

The three papers presented in this portfolio contribute to the painting of a coherent and comprehensive, albeit not totally comprehensive, picture of performance in the Australian TAFE system. This picture is coherent in that we assembled those performance indicators that complement one another in terms of coverage of several logically connected aspects of hard and soft effectiveness as well as efficiency measures. It is comprehensive (but not totally — see limitations of this study below) in that in our study we covered most aspects of TAFE performance that can be analysed using publicly available data. The purpose here was that policymakers, administrators, stakeholders and otherwise interested persons who wish to explore TAFE performance may be able to replicate or extend our analysis with easily obtainable data.

The portfolio presented here can thus be seen as bringing together a number of aspects of performance under one umbrella. On one hand we determined two discrete efficiency measures that essentially deal with the efficient use of public funds. On the other hand we developed indicators capturing hard and soft effectiveness measures. While efficiency analysis dealt with the problem how resources are transformed from inputs to outputs and outcomes, the effectiveness component asked about the quality of these outcomes. While we have shown there is not necessarily a strong relationship between efficiency and effectiveness measures, it is clear the two concepts complement each other and are useful to evaluate the various aspects of institutional and system performance. Schematically, this approach could be conceptualised in the form of a logic model. Such logical frameworks approaches have an established track record of being applied to manage and measure performance and often serve as a foundation for planning and evaluation (Herranz, 2010). In figure 5.1

we can see a graphical depiction of the conceptual relationship between inputs, outputs and outcomes of the TAFE system and how effectiveness and efficiency relate to these three categories and contribute to a holistic assessment of performance. Efficiency evaluates the quantitative nature of the relationship between inputs (resources) and outputs, whereas the evaluation of effectiveness is concerned with the qualitative aspect of the outcomes of TAFE training. In this portfolio we amalgamate these aspects of performance.

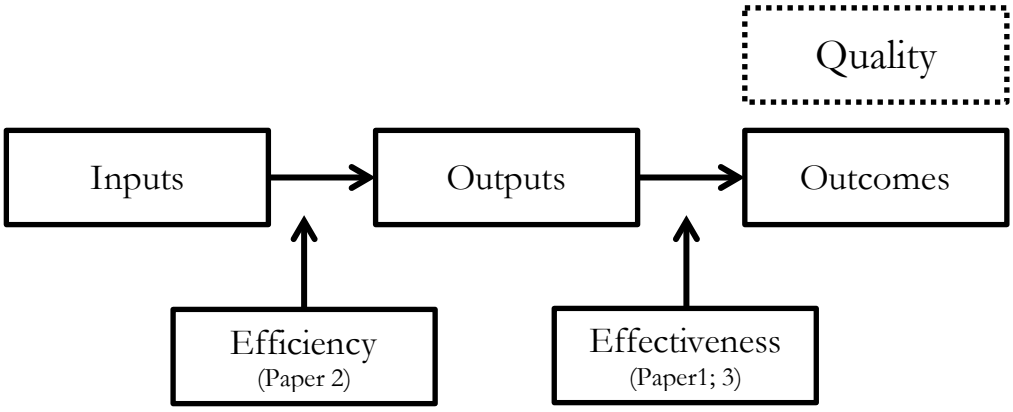

Figure 5.1 Logic model of TAFE performance measurement

The studies reported in this portfolio are the result of quantitative empirical research. As such we have gathered an array of publicly available data from a number of different sources, have formulated hypotheses and then tested these hypotheses and drawn inferences from the results. While this approach has produced tangible and, in practice, usable performance measurements that can be employed for the purpose of institutional comparisons, we also aimed to develop and present 'templates' (for example, methodologies) that can be applied by others intending to undertake such evaluations. We have employed a micro-econometric approach in this portfolio via the application of methods and data analysis pertaining to students and educational institutions. Such a methodology is ordinarily used not only in every area of micro-economics but also in related disciplines such as social sciences or political science (Cameron & Trivedi, 2005). Our approach is manifest in portfolio paper

one where we studied the impact of demographic and educational circumstances on the effectiveness on TAFE institutions. Using this approach we could develop a framework of effectiveness indicators that enables inter-institutional comparison. This framework could also be used to estimate which changes in inputs will lead to what extent of changes in effectiveness. The micro-econometric approach in our second portfolio paper considered the relationship between the inputs and outputs of TAFE institutes and how this relationship affects institutional efficiency. The third portfolio paper considered a classical microeconomic theme: 'What is the payoff to a particular action by particular actors?'. In our study this meant the benefit students may receive if they complete their courses compared with those who do not complete. The micro-econometric approach was applied in all three portfolio papers as it is such a powerful tool to examine micro-economic questions of this nature. All three portfolio papers have in common that they examine a specific outcome and, indirectly, what will happen to that outcome if certain parameters change. This is the microeconomic approach that can move debates away from the normative and toward the positive, or 'decreasing the amount of heat, while increasing the amount of light' as Levy (1995) called it. Approaching our research from a micro-economics angle lets us analyse the effects of the changes in our parameters (educational, demographic, institutional etc.) on the outcomes under investigation, be it effectiveness in the form of student satisfaction, labour market outcomes, efficiency in the form of institutional teaching load or pay offs to completion across the VET sector. We were thus able to determine the likely consequences of changes in our input variables.

5.2 Practical implications

The main benefit of the research presented in this portfolio is that new methods have been proposed enabling the assessment of institutional

effectiveness and efficiency in the publicly funded vocational education sector in Australia. The results of such assessments can then be used in the benchmarking of institutions and, if necessary, the creation of policy interventions to address possible deficiencies uncovered in the process. Furthermore, one of the declared intentions for this portfolio was to develop approaches and indicators for the measurement of efficiency and effectiveness in the Australian VET sector that are not only relevant and methodologically sound but also readily applicable in a practical environment. With this overarching goal in mind we focussed on the development of techniques that take advantage of existing data sources (or data that become routinely available on a regular basis). While currently available administrative and survey data collections have some undeniable shortcomings, such as missing data, survey bias and other attributes that complicate the comparability between institutions, we have endeavoured to incorporate steps into our methodologies that help to overcome these shortcomings. The obvious benefit of using available data with adjustments for existing limitations is that the methodologies developed here can be applied in a practical environment at very low cost and with minimal additional resources. Furthermore, application of the proposed techniques would not only be limited to central policymakers, but can be applied by institutions themselves and, in fact, anyone who desires to do so.

The research methodologies presented in papers one and two of this portfolio represent mostly tools that enable inter-institutional benchmarking and comparisons. The focus here was on providing researchers and policymakers with means permitting such comparisons of institutions that may differ substantially in respect to educational profile and student demographics. It could thus be suggested that these methods lend themselves predominantly toward the application at a central policymaker level. At this level it is most likely that information about differences in performance can be translated into

concrete funding and intervention measures. Information about relative performance compared with other providers is clearly of benefit to institutions too, so that the quantitative methods proposed in this portfolio may be equally applied by institutional researchers and planning units at individual TAFE institutes. This was one of the main reasons we aimed for the performance indicators introduced in this portfolio to rely on publicly available data, as it would subsequently enable any stakeholder (such as individual institutions or state governing bodies) to assess performance independently without relying on central agencies to release such information. The research published in this portfolio also contributes to the policy directive of greater transparency in the VET sector (Transparency Agenda in Vocational Education, 2012).

In our study we have observed some clear inefficiencies in the Australian TAFE system. These inefficiencies were mainly related to the degree of remoteness and student characteristics, both of which could be seen as exogenous to the TAFEs themselves. For example, the least efficient TAFE institutes were more likely to be found in remote locations, had a higher percentage of males, and a larger proportion of individuals from non-English speaking backgrounds. We speculate these inefficiencies were driven by a combination of interrelated factors, including geographic location, available infrastructure and the absence of occupational diversity of graduates. We also demonstrated that different types of efficiency are not necessarily linearly related with one another and that additional types of efficiencies could be defined and estimated. For policy makers it is therefore necessary to take a multi-dimensional approach that takes into account the various aspects of different methods to the concept of efficiency when making policy decisions.

The TAFE sector in Australia has a long history of providing education to low SES students, to students in regional areas and to students with low prior academic achievement. Measures of efficiency must take into account a

variety of internal and external outcome measures in determining the efficiency of the institutions. The wider, but related, notion of institutional effectiveness would integrate questions of economic efficiency and also consider the next best alternative at the student, community and national level that would flow from the absence of the TAFE institutions. In many instances, the alternative to TAFE education and a skilled job may well by unemployment, social and economic marginalisation and welfare dependency.

This emphasizes that in the efficiency analysis of educational institutions it is necessary that any efficiency model needs to be specified with a clear purpose in respect to which particular aspect of institutional efficiency is going to be investigated, and these efficiency aspects need to be considered in a wide context that investigates both internal operational elements and also the social benefits that may flow from the upskilling of an archetypical TAFE student.

While the performance indicators introduced in portfolio papers one and two dealt with the measurement of performance at the institutional level, portfolio paper three investigated the specific issue of completion in vocational education at a more systemic level. However, one aim here was also the practical applicability of the methodology at an individual or institutional level. The application of the methodology to determine the 'completion deficit' could be considered as an example of the practical utility of the approach developed. Here administrators of individual institutions could use available demographic, institutional and educational enrolment data to calculate a theoretical 'completion deficit' score for each individual student at the point of enrolment. All the necessary information is available before the prospective student commences studies. Education providers could thus be in a position to identify individuals with a high probability of not completing their studies at the outset of their programs and thus be in a position to implement

special intervention measures aimed at students likely to discontinue their studies before completing a qualification. Such intervention measures could include efforts to increase student motivation, frequent follow up, academic advising, supplementary instruction and others.

5.3 Limitations of this portfolio of research

As in other research, there are a number of limitations to the research presented in this portfolio. While we attempted to undertake a comprehensive analysis of research questions posed, some issues were impossible to resolve in the confines of this study and were beyond the control of the author.

The research includes only publicly funded vocational education providers: This study only includes publicly funded VET providers. The analysis in the first two portfolio papers is based predominantly on TAFE institutes. This is largely due to the availability of data. While the Student Outcome Survey (SOS) does collect data from private and public providers of VET education, sample sizes and survey design are only sufficient to make statistically relevant statements about public institutions (private provider data is ordinarily aggregated to the state level). Any of the results presented in portfolio papers one and two can therefore not be extrapolated to the private Australian VET sector. The third portfolio paper includes, in addition to the TAFE population, those students who enrolled in vocational education programs at private providers, as long as these providers also receive public funding. The reason for this extended population in paper three is the availability of this kind of data from private providers. As portfolio paper three did not aim to facilitate institutional comparisons we could include students attending private providers while increasing the statistical power of our analysis due to the increased student population.

Efficiency is not absolute but relative:

Our portfolio paper two on efficiency could only make substantive conclusions about the efficiency of TAFE institutes relative to each other. The methodologies proposed (Data Envelopment Analysis as well as Stochastic Frontier Analysis) are insufficient to determine absolute efficiency, even though some institutes will be assigned an (theoretically perfect) efficiency of 1. Furthermore, other circumstances, which may not necessarily be observed, may impact on efficiency. For instance, a hypothetical institute may incur significant additional cost for maintenance of a historical campus. This may have a detrimental impact on its estimated efficiency due to its disproportionally higher capital costs.

Data quality:

In this portfolio we relied mostly on survey and administrative data with smaller amounts of data obtained by personal requests and from institutional annual reports. Specifically survey and administrative data are well known to be error prone. Respondents may not tell the truth, may make errors when filling in forms, may accidentally enter responses in the wrong space or omit data. Equally, administrative data may be erroneous due to transcription errors, faulty formatting, omissions or other data errors. While we took great care when preparing our analysis, the possibility cannot be excluded that a small percentage of data contains errors.

Data availability:

One of the declared aims of this portfolio was to answer the research questions posed with data that was either freely available or relatively easily obtainable (for example, via informal data requests etc.). The main reason for this aim was to underscore the practical relevance of this research as easily obtainable data may be beneficial in promoting the practical usage of the methodologies proposed herein. While we consider the reliance on such data in this portfolio of great benefit for the aforementioned reason, it could also be

considered to be a disadvantage as this approach restricted the scope of variables that we could incorporate into our analysis.

Survey issues/bias:

Much of the analysis in this portfolio was performed using survey data, mostly from the Student Outcome Survey. While this is a very large survey (300,000+ questionnaires) and therefore provides sizeable statistical power, it has all the inherent problems of survey research, such as sampling error, non-sampling error, survey bias, non-response bias, data errors etc. However, we attempted to alleviate several of these survey shortcomings by incorporating some methods in our analysis designed to counter the effect of such problems.

Unobserved variables:

In this portfolio we attempted to answer the research questions posed using all the data at our disposal. It was, however, clear that there were often other, unobserved, traits which may have explanatory power. The analysis cannot account for these variables. It should thus be noted that the performance measures developed in this portfolio should be interpreted as 'indicators' only. This means that the performance measure under consideration should be thought of as valid if 'all other things are equal'. This, in reality, may not always be the case.

Causality:

In this research, we determined quantitative relationships between dependent and independent variables. These findings are by nature correlational, and while there may be causative relationships, alternative explanations cannot necessarily be ruled out. While there are approaches such as counterfactual theory that can be used to assess causality, to show causality in an empirical context it would have been necessary to conduct randomized controlled experiments, an endeavour that is clearly beyond the scope of this portfolio.

Timeliness:

While it was one aim of this research to develop methodologies that can be used by researchers irrespective of any given timeframe, we also intended to provide a snapshot of measured performance in the contemporary VET environment. It was therefore our objective to use the latest available data to give an accurate and current assessment of performance in the ATS. Due to data issues (such as the alternating extent of the Student Outcome Survey between odd and even years, and the concurrent availability of data from the Student Intentions Survey) and the time passed while compiling this research portfolio most research results presented herein is representative of the year 2011. While this represents a sizeable lag between the time of data collection and the publication of the results of our research, we believe that, by conventional academic standards, the magnitude of this time lag can still be considered satisfactory.

Political environment:

Another limitation stems from the political climate surrounding tertiary education and how this affects the construction and availability of certain measures and the relative emphasis on efficiency or effectiveness in performance reporting. Furthermore, the often obscure delineation between federal and state government involvement in the VET sector as well as rapid changes in the policy environment may lead to increased abundance of usable data, but conversely may also result in fewer data items available to apply the methodologies developed in this research.

Limitation of the positivist approach used in this portfolio:

Finally, the positivist analytical framework employed in this portfolio may impose limitations by focusing solely on quantitative indicators. The absence of components that are hard to quantify (for instance, human emotional perspectives and reactions) and a certain inflexibility to deviate from a

structured, dispassionate approach may potentially undermine the strength of the positivist approach. There may be value in adopting a more systemic mixed assumptions/methods approach in order to gain insights into the perspectives of a range of stakeholders in the TAFE educational process. Such an approach might be particularly useful when dealing with the 'human element' at the institutional level.

5.4 Pathways for future research

In this research portfolio we analysed a number of key aspects of performance in the publicly funded Australian vocational education system by developing methodologies to create performance indicators and the application of such methods. Our efforts were constrained by a number of impediments, as outlined in the previous paragraph. Naturally, the first and obvious path for future research building on this portfolio would be to attempt to overcome such impediments. Generally this might include endeavours to acquire more up-to-date data, the introduction of additional measures to improve data quality, further elimination of biases, improvement of the quality of the created performance indicators or simplifying the methodology in an effort to make the research more accessible to non-technical audiences. In addition, further research may and should be taking advantage of the unique student identifier (USI) that will be introduced in 2015. Once introduced and in force for some period of time, the USI will bring clarity to a number of open questions in the VET research discipline, most importantly it will enable the accurate estimation of completion rates. In this portfolio we employed completion rates in papers two and three and estimated those using specific assumptions. The presence of the USI in future years signifies that additional research could be conducted to validate the findings in this portfolio and expand on research results.

The aforementioned possibilities for future research are of a type of general nature and apply to more than one of the research papers presented in this portfolio. However, there are also a number of quite specific research avenues pertaining to the three portfolio papers that might be worthwhile to explore in future research.

In portfolio paper one, we developed seven specific institutional performance indicators using mixed effect models. The indicators were adjusted for

potential non-response bias inherent in the survey data. We then created clusters to assess the performance profiles of all indicators against the clusters. Finally we created a summary performance indicator using a Euclidean distance algorithm. The techniques developed in this paper could be further expanded by the purposive collection of additional data and the creation of additional indicators that are relevant to the changing VET environment. Furthermore, such additional performance indicators could be designed to be more reflective of the programmatic nature of VET funding. In order to conduct future research it will be necessary to investigate what other information may be necessary to accomplish this. There are currently a number of policy initiatives in progress which may provide a source of data for this purpose. Such policies include VET FEE HELP, a policy designed to assist eligible students studying higher-level VET qualifications (such as diplomas and advanced diplomas) to pay their tuition fees (VFH 2014) and the effects of the Victorian Entitlement Model, which provides for an individual entitlement to a subsidised training place in the State of Victoria (Leung et al., 2014).

Our second portfolio paper assessed the technical efficiencies of TAFE institutes using two different definitions of TAFE efficiency and two competing methods to determine these efficiencies. We compared the results from both definitions and methods and then applied discriminant analysis to develop a typology of efficient institutes. Recent developments in the government-funded VET sector have seen the amalgamation of a number of institutes (ABC, 2014a; 2014b; *Brisbane Times,* 2013). As the number of available institutes for efficiency analysis was already somewhat low at the time of conducting the research in this portfolio, any further reduction may necessitate the re-evaluation of the research methodology applied. This is particularly true for the component that relied on parametric analysis. Further research could thus investigate alternative methodologies that facilitate efficiency analysis in the VET sector. Another avenue for additional research building on this paper

would be the application of a meta-frontier framework which would allow us to estimate the performance within and between certain categories. This could, for example, include analysis of private versus public or rural versus urban institutions in respect to their efficiency. This portfolio has only given limited consideration to the effectiveness—efficiency nexus, which represents an additional avenue for further research. Particularly interesting would be to explore whether efficiency may in some respect have a deleterious effect on effectiveness, as could be inferred from the negative relationship between teaching efficiency and effectiveness as seen in table 5.1.

Finally, in paper three we investigated the topic of course completions in vocational education at the system level. Here, we used the probabilities that an individual student would complete his studies to determine the payoff to completion. Chi squared automated interaction detection was then used to aggregate the payoffs into distinct student groups of demographic or educational interest. We concluded the third portfolio paper by examining the relationship between the probability to complete and the probability to intend to complete and by using these two measures to define and estimate the 'completion deficit', a measure that allows identification of potential dropout students at the point of their enrolment. One central theme in portfolio paper three was the estimation of 'completion benefits' for various outcomes. In this paper we approached the topic of individual completion benefits without taking the cost on the side of the student into account. To further refine this research it would be beneficial to include these costs in our analysis, with the aim of determining net benefits. In this portfolio it was beyond the scope of the analysis, given that data on such costs would be difficult to obtain. Future research could be design to overcome this limitation. The research presented in this portfolio paper could also be extended by examining possible interactions between institutional level and student level. Furthermore, as already pointed out earlier, the introduction of the unique student identifier

will have the strongest bearing on potential research opportunities emanating from this paper as it will enable a more accurate estimation of completion rates.

In addition to the aforementioned pathways for further research, another possible additional approach to enhance the research presented in this portfolio could encompass a modified methodological paradigm. Such refined framework could include post-positivism and thus offer greater scope for paradigm shift based on the uncovering of new evidence.

5.5 Conclusion

In this doctoral portfolio we have attempted to draw a picture of contemporary performance measurement issues in the vocational education sector in Australia. Our aim was to develop methodologies that facilitate the accurate assessment of several performance dimensions, create indicators that are easily interpretable and can be utilised in performance measurement and to estimate these performance measures in order to create a snapshot of contemporary performance using recent data. Most of the data we employed could be obtained from the public domain, so that it will be uncomplicated for others to replicate the methods developed here and apply them to data released in the future. This portfolio employed a broad array of statistical techniques with the intention of making a contribution to the application of quantitative methods in the VET research field.

The vocational education system in Australia is a fast paced environment and demographic, societal and policy changes, along with administrative changes such as institute mergers, plus the friction created by the sometimes blurred demarcation between state and federal policy involvement in the VET sector, provide for frequent and difficult-to-predict systemic changes. Any comprehensive analysis of the Australian VET system must therefore always be

viewed in the context and timeframe in which it was performed. In this portfolio we aimed to develop methodologies that can mostly be applied irrespective of administrative changes. It is hoped that some of the research presented here may be of use in an applied sense by policymakers and researchers.

6 References

Abbott, M & Doucouliagos, C (1998). *DEA analysis of the efficiency of Victorian technical and further education institutes*. Working paper. Burwood, Victoria, Deakin University.

—— (2002). 'A data envelopment analysis of the efficiency of Victorian TAFE institutes'. *Australian Economic Review*, vol.35, no.1, pp.55—69.

—— (2003). 'The efficiency of Australian universities: a data envelopment analysis'. *Economics of Education Review*, vol.22, no.1, pp.89—97.

—— (2009). 'Competition and efficiency: overseas students and technical efficiency in Australian and New Zealand universities'. *Education Economics*, vol.17, no.1, pp.31—57.

ABC (2014a). *AEU: Gippsland TAFE merger result of government funding cuts*. News report, retrieved 12 February 2015 from <http://www.abc.net.au/local/audio/2014/04/16/3987070.htm>.

—— (2014b). *Bendigo Kangan TAFE to cut hundreds of jobs, union says*. News report, retrieved 12 February 2015 from <http://www.abc.net.au/news/2014-12-05/bendigo-kangan-tafe-to-cut-hundreds-of-jobs-union-claims/5946496>.

ABS (Australian Bureau of Statistics) (2011). *Census 2011*. Retrieved May 2013 from <http://www.abs.gov.au/websitedbs/censushome.nsf/home/data?opendocument#from-banner=LN>.

—— (2013). *Socio-economic indexes for areas*. Retrieved June 2013 from <http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa>.

—— (2013a). Australian Standard Classification of Education. Retrieved June 2013 from <http://www.abs.gov.au/ausstats/abs@.nsf/0/E7779A9FD5C8D846CA256AAF001FCA5C?opendocument>.

—— (2013b). Australian Standard Geographical Classification. Retrieved June 2013 from <http://www.abs.gov.au/Ausstats/abs@.nsf/0/0D204FD3DCD90564CA256F19001303A2>.

ACER (2011). *University retention and completion rates have improved.* Media release, retrieved July 1 2014, <http://www.acer.edu.au/media/article/university-retention-and-completion-rates-have-improved>.

Aigner, D, Lovell, CA & Schmidt, P (1977). 'Formulation and estimation of stochastic frontier production function models'. *Journal of Econometrics*, vol.6, no.1, pp.21—37.

Anderson, D (1998). 'The Metamorphosis of TAFE', *Australian and New Zealand Journal of Vocational Education Research*, vol.6, issue 2, 1998.

Anfara, V & Mertz, N (2006). 'Introduction', in *Theoretical frameworks in qualitative research*. pp.xiii-xxxii . Thousand Oaks, CA. Sage.

ANR (2011). *Annual National Report of the Australian Vocational Education and Training System 2011,* Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education, Canberra. Retrieved October 2013 from <http://www.innovation.gov.au/skills/ResourcesAndPublications/Documents/ANRVET2011.pdf>.

ANTA (Australian National Training Authority) (1994). '*Key performance measures for vocational education and training, 1994'*. Adelaide. ANTA.

—— (1996). '*Key performance measures for vocational education and training, 1996'*. Adelaide. ANTA.

AQF (Australian Qualification Framework) (2013). Retrieved June 2013 from <http://www.aqf.edu.au/AbouttheAQF/AQFQualifications/Qualifications/tabid/194/Default.aspx>.

ASQA (Australian Skills Quality Authority) (2012). *General direction: quality indicators*. Retrieved September 2013 from <http://www.asqa.gov.au/verve/_resources/General_direction_-_Quality_indicators_24_April_2012.pdf#search=indicators>.

AVETMISS (2011). *AVETMISS 6.1 for VET providers: what's new and why*? Adelaide. NCVER. Retrieved 4 March 2013 from <http://www.ncver.edu.au/publications/2401.html>.

Azemikhah, H (2009). 'Aiming at the higher completion rates in VET'. Paper presented at 12th Australian Vocational Education and Training Research Association (AVETRA) Conference. Sydney. Retrieved May 2013 from <http://www.avetra.org.au/papers-2009/papers/79.00.pdf>.

Baecke, P & Van den Poel, D (2012). 'Including the salesperson effect in purchasing behavior models using PROC GLIMMIX'. *SAS Forum Statistics and Data Analysis*. SAS Inc., Orlando, Florida

Banker, RD (1989). 'Econometric estimation and data envelopment analysis'. *Research in Government and Nonprofit Accounting*, vol.5, pp.231—43.

—— (1996). 'Hypothesis tests using data envelopment analysis'. *Journal of Productivity Analysis*, vol.7, no.2—3, pp.139—59.

Banker, RD, Charnes, A & Cooper, WW (1984). 'Some models for estimating technical and scale inefficiencies in data envelopment analysis'. *Management Science*, vol.30, no.9, pp.1078—92.

Banker, RD & Morey, RC (1986). 'The use of categorical variables in data envelopment analysis'. *Management Science*, vol.32, no.12, pp.1613—27.

Beccalli, E, Casu, B & Girardone, C (2006). 'Efficiency and stock performance in European banking'. *Journal of Business Finance & Accounting*, vol.33, no.1—2, pp.245—62.

Bednarz, A (2012). *Lifting the lid on completion rates in the VET sector: how they are defined and derived*. NCVER. Adelaide.

Biørn, E, Hagen, TP, Iversen, T & Magnussen, J (2003). 'The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA efficiency scores 1992—2000'. *Health Care Management Science*, vol.6, no.4, pp.271—83.

Bjurek, H, Kjulin, U & Gustafsson, B (1992). 'Efficiency, productivity and determinants of inefficiency at public day care centers in Sweden'. *The Scandinavian Journal of Economics*, pp.S173—S187.

Bond, TG & Fox, CM (2013). *Applying the Rasch model: Fundamental Measurement in the Human Sciences*. Psychology Press.

Bontempo, J & Morgan P (2001). *TAFE student survey validation — measuring student perceptions of teaching, assessment and learning'*. Western Australia. The Department of Education.

Borland, J (2002). *New estimates of the private rate of return to university education in Australia*. Melbourne. Melbourne Institute of Applied Economic and Social Research.

Bradley, S, Johnes, G & Millington, J (2001). 'The effect of competition on the efficiency of secondary schools in England'. *European Journal of Operational Research*, vol.135, no.3, pp.545—68.

Bradley, S, Draca, M & Green, C (2004). 'School performance in Australia: is there a role for quasi-markets?'. *Australian Economic Review*, vol.37, no.3, pp.271—86.

Brisbane Times (2013). 'TAFE and university merge in Queensland first'. News report. Retrieved 12 February 2015 from <http://www.brisbanetimes.com.au/ queensland/tafe-and-university-merge-in-queensland-first-20130911-2tjpy.html>.

Butler, TW & Johnson, WW (1997). 'Efficiency evaluation of Michigan prisons using data envelopment analysis'. *Criminal Justice Review*, vol.22, no.1, pp.1—15.

Cameron, AC & Trivedi, PK (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Cameron, K (1978). 'Measuring organizational effectiveness in institutions of higher education'. *Administrative Science Quarterly*, pp.604—32.

Cameron, R (2010). 'Mixed methods in VET research: usage and quality'. *International Journal of Training Research*, vol.8, no.1, pp.25—39.

Carrington, R, Coelli, T & Rao, P (2004). 'Measuring the performance of Australian universities: conceptual issues and initial results'. Conference presentation. July 2004. Brisbane. Asia-Pacific Productivity Conference.

—— (2005). 'The performance of Australian universities: conceptual issues and preliminary results'. *Economic Papers*, vol.24, no.2, pp.145—63.

Chappel, C (2003). 'Researching vocational education and training: where to from here?'. *Journal of Vocational Education and Training*, vol.55, no.1, pp.21—32.

Charnes, A, Cooper, WW & Rhodes, E (1978). 'Measuring the efficiency of decision making units'. *European Journal of Operational Research*, vol.2, no.6, pp.429—44.

Chen, CF (2007). 'Applying the stochastic frontier approach to measure hotel managerial efficiency in Taiwan'. *Tourism Management*, vol.28, no.3, pp.696—702.

Churchill Jr, GA, Ford, NM & Walker Jr, OC (1974). 'Measuring the job satisfaction of industrial salesmen'. *Journal of Marketing Research*, pp.254—60.

Cobb, CW & Douglas, PH (1928). 'A theory of production'. *The American Economic Review*, vol.18, no.1, pp.139—65.

Cohen, L, Manion, L & Morrison, K (2007). *Research methods in education*, London & New York. Routledge.

Cohen, AM & Brawer, FB (1996). *The American community college*. San Francisco. Jossey-Bass Publishers.

Cullinane, K, Song, DW & Gray, R (2002). 'A stochastic frontier model of the efficiency of major container terminals in Asia: assessing the influence of administrative and ownership structures. *Transportation research part A: policy and practice*, vol.36, no.8, pp.743—62.

Cullinane, K, Wang, TF, Song, DW & Ji, P (2006). 'The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis. *Transportation research part A: policy and practice*, vol.40, no.4, pp.354—74.

Cullinane, K & Song, DW (2006). 'Estimating the relative efficiency of European container ports: a stochastic frontier analysis'. *Research in Transportation Economics*, vol.16, pp.85—115.

Curtis, D (2010), 'Evaluating institutional performance indicators in VET', (unpublished). Adelaide. National Centre for Vocational Education Research.

Daghbashyan, Z (2011). *The economic efficiency of Swedish higher education institutions*. CESIS, KTH Royal Institute of Technology Report.

DEEWR (Department of Education, Employment, and Workplace Relations) (2011). *Australian regional labour markets publication.* Retrieved September 2013 from <http://deewr.gov.au/australian-regional-labour-markets-publication>.

—— (2011a). *Annual national report of the Australian VET sector: 2010*. Canberra. Government of Australia.

De Leeuw, ED & De Heer, W (2002). 'Trends in household survey nonresponse: a longitudinal and international comparison'. *Survey Nonresponse*, pp.41—54.

DET (Department of Education) (2014). *Total VET activity and the unique student identifier*. Retrieved 4 March 2013 from <http://www.det.act.gov.au/training/skills_reform/unique_student_identifier>.

Deprins, D, Simar, L & Tulkens, H (1984). 'Measuring labor inefficiency in post offices. The performance of public enterprises'. *Concepts and measurements*, eds M Marchand, P Pestieau and H Iulkens, pp.243—67. North-Holland, Amsterdam.

Deville, JC, Särndal, CE & Sautory, O (1993). 'Generalized raking procedures in survey sampling'. *Journal of the American Statistical Association*, vol.88, no.423, pp.1013—20.

Edwards, JE & Waters, LK (1982). 'Involvement, ability, performance and satisfaction as predictors of college attrition'. *Educational and Psychological Measurement*, vol.42, pp.1149—52.

Farrell, MJ (1957). 'The measurement of productive efficiency'. *Journal of the Royal Statistical Society*. Series A (General), vol.120, no.3, pp.253—90.

Fatulescu, P (2013). 'A DEA-SFA comparison of the impact of ICT's utilization'. *European Scientific Journal*, vol.9, no.21.

Fieger, P (2010). *An investigation of TAFE efficiency'*. Adelaide. National Centre for Vocational Education Research. Retrieved July 2013 from <http://avetra.org.au/wp-content/uploads/2010/04/42.00-Peter-Fieger.pdf>.

Fieger, P, Karmel, T & Stanwick, J (2010). *An investigation of TAFE efficiency.* Adelaide. National Centre for Vocational Education Research. Retrieved July 2013 from <http://www.ncver.edu.au/publications/2290.html>.

Fischer, GH & Molenaar, IW (eds) (1995). *Rasch models: foundations, recent developments, and applications*. Springer.

Forrester, JW (1994). 'System dynamics, systems thinking, and soft OR'. *System Dynamics Review*, vol.10, no.2—3, pp.245—56.

Fraser, I & Cordina, D (1999). 'An application of data envelopment analysis to irrigated dairy farms in Northern Victoria, Australia'. *Agricultural Systems*, vol.59, no.3, pp.267—82.

Friedman, M (1979). 'Truth and confirmation'. *The Journal of Philosophy,* vol.76, no.7, pp.361—82.

Goozee, G (2001), *The development of TAFE in Australia.* 3rd edition. Adelaide. National Centre for Vocational Education Research, pp.1—44.

Grice, James W (2001). 'Computing and evaluating factor scores'. *Psychological Methods*, vol.6, no.4, pp.430—50.

Guthrie, H (1991). *Performance indicators revisited.* Adelaide. TAFE National Centre for Research and Development.

Halfpenny, P (1992). *Positivism and sociology: explaining social life*. London. Allen and Unwin.

Hanke, M & Leopoldseder, T (1998). 'Comparing the efficiency of Austrian universities: a data envelopment analysis application'. *Tertiary Education & Management*, vol.4, no.3, pp.191—7.

Harris, R & Clayton B (2010). 'Impact in vocational education and training research: the case of the Australian VET Research Consortium. *International Journal of Training Research*, vol.8, pp.6—24.

Hartigan, JA (1985). 'Statistical theory in clustering'. *Journal of Classification*, vol.2, no.1, pp.63—76.

Hashimoto, K & Cohn, E (1997). 'Economies of scale and scope in Japanese private universities'. *Education Economics*, vol.5, no.2, pp.107—15.

Hauser, RM & Warren, JR (1997). 'Socioeconomic indexes for occupations: a review, update, and critique'. *Sociological Methodology*, vol.27, no.1, pp.177—298.

Heckman, JJ (1979). 'Sample selection bias as a specification error'. *Econometrica*, vol.47, no.1, pp.153—61.

Hedecker, D (2003). 'A mixed effects multinomial logistic regression model', *Statistics in Medicine*, vol.22, pp.1433—46.

Herault, N, Zakirova, R & Buddelmeyer, H (2011). *The effect of VET completion on the wages of young people*. Adelaide. NCVER.

Herranz Jr, J (2010). 'The logic model as a tool for developing a network performance measurement system'. *Public Performance & Management Review,* vol.34, no.1, pp.56—80.

Higgins, JC (1989). 'Performance measurement in universities'. *European Journal of Operational Research*, vol.38, no.3, pp.358—68.

Horn, JL (1965). 'A rationale and test for the number of factors in factor analysis'. *Psychometrika*, vol.30*,* pp.179—85.

Horne, J & Hu, B (2008). 'Estimation of cost efficiency of Australian universities'. *Mathematics and Computers in Simulation*, vol.78, no.2, pp.266—75.

Homs, O (2007). *Euromed observatory function indicators for the governance of vocational training systems*. Torino. European Training Foundation.

Hosmer, DW & Lemeshow, S (2000). 'Model-building strategies and methods for logistic regression'. *Applied Logistic Regression, second edition*, pp.91—142.

Hunter, B (1996). *Indigenous Australians and the socioeconomic status of urban neighbourhoods*. Canberra. Australian National University, Centre for Aboriginal Economic Policy Research.

Izadi, H, Johnes, G, Oskrochi, R & Crouchley, R (2002). 'Stochastic frontier estimation of a CES cost function: the case of higher education in Britain'. *Economics of Education Review*, vol.21, no.1, pp.63—71.

JMetrik, (2013). An open source computer program for psychometric analysis. Retrieved in May 2013 from <http://www.itemanalysis.com/index.php>.

James, R (2001). 'Participation disadvantage in Australian higher education: an analysis of some effects of geographical location and socioeconomic status'. *Higher Education*, vol.42, no.4, pp.455—72.

Kalaitzandonakes, NG & Dunn, EG (1995). 'Technical efficiency, managerial ability and farmer education in Guatemalan corn production: a latent variable analysis'. *Agricultural and Resource Economics Review*, vol.24, no.1.

Kanep, H (2004). *Assessing the efficiency of Estonian secondary schools: an application of data envelopment analysis*. Working paper.

Kangan, M (1974). *TAFE in Australia: report on the needs of technical and further education (Kangan Report).* Australian Committee on Technical and Further Education (ACOTAFE). Canberra. AGPS.

Karlaftis, MG (2004). 'A DEA approach for evaluating the efficiency and effectiveness of urban transit systems'. *European Journal of Operational Research*, vol.152, no.2, pp.354—64.

Karmel, T (2011). 'Year 12 completions and youth transitions: research overview'. *Longitudinal Surveys of Australian Youth, research overview: research report 56*. Adelaide. NCVER.

Karmel, T, Fieger, P, Blomberg, D & Loveder, P (2013). *Performance indicators in VET.* Adelaide. NCVER.

Karmel, T & Mlotkowski, P (2010). *The impact of wages on the probability of completing an apprenticeship or traineeship.* Adelaide. NCVER.

Karmel, T & Nguyen, N (2006). *The value of completing a vocational education and training qualification.* Adelaide. NCVER.

Kass, G (1980). 'An exploratory technique for investigating large quantities of categorical data'. *Applied Statistics*, *vol.*29, no.2, pp.119—27.

Kempkes, G & Pohl, C (2010). 'The efficiency of German universities — some evidence from non-parametric and parametric methods'. *Applied Economics*, vol.42, no.16, pp.2063—79.

Kirjavainen, T & Loikkanent, HA (1998). 'Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis'. *Economics of Education Review*, vol.17, no.4, pp.377—94.

Klassen, A, Miller, A, Anderson, N, Shen, J, Schiariti, V & O'Donnell, M (2010). 'Performance measurement and improvement frameworks in health, education and social services systems: a systematic review'. *International Journal for Quality in Health Care*, vol.22, no.1, pp.44—69.

Kraft, E, Hofler, R & Payne, J (2006). 'Privatization, foreign bank entry and bank efficiency in Croatia: a Fourier-flexible function stochastic cost frontier analysis'. *Applied Economics*, vol.38, no.17, pp.2075—88.

Kruskal, WH & Wallis, WA (1952). 'Use of ranks in one-criterion variance analysis'. *Journal of the American Statistical Association*, vol.47, no.260, pp.583—621.

Laband, DN & Lentz, BF (2003). 'New estimates of economies of scale and scope in higher education'. *Southern Economic Journal*, pp.172—83.

Lee, C & Ji, YB (2009, December). 'Data envelopment analysis in stata', in *Computer program presented at Stata Conference.* Washington DC.

Lenton, P (2008). 'The cost structure of higher education in further education colleges in England'. *Economics of Education Review*, vol.27, no.4, pp.471—82.

Leow, C, Marcus, S, Zanutto, E & Boruch, R (2004). 'Effects of advanced course-taking on math and science achievement: addressing selection bias using propensity scores'. *American Journal of Evaluation*, vol.25, no.4, pp.461—78.

Leung, F, McVicar, D, Polidano, C & Zhang, R (2014). *Early impacts of the Victorian Training Guarantee on VET enrolments and graduate outcomes*. Research report. Melbourne University. Retrieved in February 2015 from <https://melbourneinstitute.com/downloads/labour/Early-impacts-of-the-VTG-2724.pdf>.

Levy, JM (1995). *Essential microeconomics for public policy analysis*. New York. Praeger.

Long, M, Carpenter, P & Hayden, M (1995). *Graduating from higher education,* DEET. Canberra. AGPS.

Long, M & Shah, C (2008). *Private returns to vocational education and training qualifications. a national vocational education and training research and evaluation program report*. Adelaide. NCVER.

Loveder, P & Guthrie, H (2008). *Genesis of VET research: case study of Australia*. UNESCO Handbook of Technical and Vocational Education and Training Research, eds F Rauner & R Maclean. Springer Press.

Mark, K & Karmel, T (2010). *The likelihood of completing a VET qualification: a model based approach*. Technical paper. Adelaide. NCVER.

Marks, G (2007). *Completing university : characteristics and outcomes of completing and non-completing students*. Longitudinal Surveys of Australian Youth research report 51. Retrieved in August 2013 from <http://research.acer.edu.au/lsay_research/55>.

Martin, T, O'Donoghue, T & O'Neill, M (2012). *Teachers and teaching in vocational education and training institutions — reflections from Western Australia*, New York. Nova Science Publishers.

Martinez, P (2001*). Improving student retention and achievement: what do we know and what do we need to find out?*. Report. Learning and Skills Development Agency.

McInnis, C, Hartley, R, Polesel, J & Teese, R (2000). *Non-completion in vocational education and training in higher education: a literature review*. Commissioned by the Department of Education, Training and Youth Affairs.

McIntyre, D, Rogers, L & Heier, EJ (2001). 'Overview, history, and objectives of performance measurement'. *Health Care Financing Review*, vol.22, no.3, pp.7—22.

McMillan, J, Beavis, A & Jones, FL (2009). 'The AUSEI06: A new socioeconomic index for Australia'. *Journal of Sociology*, vol.45, no.2, pp.123—49.

Mills, J, Bowman, K, Crean, D & Ranshaw, D (2012). *Workforce skills development and engagement in training through skill sets: literature review*. Occasional paper. Adelaide. NCVER.

Misko, J & Halliday-Wynes, S (2009). *Tracking our success: how TAFE institutes evaluate their effectiveness and efficiency*. Adelaide. NCVER.

Mizala, A, Romaguera, P & Farren, D (2002). 'The technical efficiency of schools in Chile'. *Applied Economics*, vol.34, no.12, pp.1533—52.

Nayar, P & Ozcan, Y (2008). 'Data envelopment analysis comparison of hospital efficiency and quality', *Journal of Medical Systems*, vol.32, no.3, pp.193—9.

NCVER (National Centre for Vocational Education Research) (2002). *At a glance — vocational education and training in Australia 1991—2000*. Adelaide. NCVER.

—— (2009). *Outcomes from the Productivity Placement Program*. Adelaide. NCVER. Retrieved 11 August 2012 from <http://www.ncver.edu.au/publications/2220.html>.

—— (2010). Financial information. Retrieved 15 August 2012 from <http://www.ncver.edu.au/publications/2426.html>.

—— (2011) *'Australian vocational education and training statistics — student intentions 2011'*. Statistical report. Adelaide. NCVER.

—— (2011b). *Student Outcomes Survey*. Adelaide. NCVER. Retrieved 25 October 2012 from <http://www.ncver.edu.au/statistic/21065.html>.

—— (2011c). *Student Intentions Survey*. Adelaide. NCVER. Retrieved 12 March 2013 from <http://www.ncver.edu.au/publications/2425.html>.

—— (2011d), *Students and Courses.* Adelaide. NCVER. Retrieved 2 September 2013 from <http://www.ncver.edu.au/statistic/21053.html>.

—— (2011e), *Student Outcomes Survey, technical notes*. Adelaide. NCVER. Retrieved 1 April 2013 from <http://www.voced.edu.au/content/ngv49354>.

—— (2013). *Australian Quality Training Framework*. Retrieved 1 March 2013 from <http://www.ncver.edu.au/resources/glossary/glossary_full_record.html?query=BROWSE&entry=Australian+Quality+Training+Framework>.

—— (2014). *Data Access Policy*. Retrieved 15 August 2012 from <http://www.ncver.edu.au/content/Data_Access_and_Charging_Policy.pdf>.

—— (2014). Financial information. Retrieved 15 August 2012 from <http://www.ncver.edu.au/publications/2426.html>.

—— (2014). *Privacy Policy*. Retrieved 15 August 2012 from <http://www.ncver.edu.au/content/Privacy_Policy.pdf>.

—— (2013). *Protocols for collecting and reporting VET statistical information.* Retrieved 15 August 2012 from <http://www.ncver.edu.au/content/Protocols_for_collecting_and_reporting_VET_Statistical_Information.pdf>.

Neely, A (1999). 'The performance measurement revolution: why now and what next?'. *International Journal of Operations & Production Management*, vol.19, no.2, pp.205—28.

Newman, C & Matthews, A (2006). 'The productivity performance of Irish dairy farms 1984–2000: a multiple output distance function approach'. *Journal of Productivity Analysis*, vol.26, no.2, pp.191—205.

Nisantha, KA & Ranasinghe, MADL (2012). 'Use of stochastic production frontiers for measuring learning efficiency: evidence from a state university in Sri Lanka'. In *First International Economic Research Conference of the Sri Lanka Forum of University Economists 19—20 November 2012*. pp.137—42. Colombo. University of Colombo.

Oliveira, M & Santos, C (2005). 'Assessing school efficiency in Portugal using FDH and bootstrapping'. *Applied Economics*, vol.37, no.8, pp.957—68.

Oliver, D (2012). *Lower-level qualifications as a stepping stone for young people*. Adelaide. NCVER.

Önder, Ö, Deliktas, E & Karadag, M (2013). 'The comparison of DEA and SFA methods in the efficiency of the Turkish manufacturing industry'. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol.18, no.1.

Oren, MN & Alemdar, T (2006). 'Technical efficiency analysis of tobacco farming in Southeastern Anatolia'. *Turkish Journal of Agriculture and Forestry*, vol.30, no.2, p.165.

Pahwa, A, Feng, X & Lubkeman, D (2003). 'Performance evaluation of electric distribution utilities based on data envelopment analysis'. *Power Systems, IEEE Transactions*, vol.18, no.1, pp.400—5.

Parsons, LJ (2004). 'Measuring performance using stochastic frontier analysis: an industrial salesforce illustration'. *ISBM Report*, 6—2004.

Preston, A (1997). 'Where are we now with human capital theory in Australia'. *Economic Record*, vol.73, no.220, pp.51—78.

Procházková, J (2011). 'Efficiency of hospitals in the Czech Republic (No. 2011/02)'. PhD thesis. Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Puhani, P (2000). 'The Heckman correction for sample selection and its critique'. *Journal of Economic Surveys*, vol.14, no.1, pp.53—68.

Ragheb, MG & Griffith, CA (1982). 'The contribution of leisure participation and leisure satisfaction to life satisfaction of older persons'. *Journal of Leisure Research*, vol.14, no.4, pp.295—306.

Ramsden, P (1991). 'A performance indicator of teaching quality in higher education: The Course Experience Questionnaire'. *Studies in Higher Education*, vol.16, no.2, pp.129—50.

Rao, JNK & Scott, AJ (1981). 'The analysis of categorical data from complex surveys: Chi-squared tests for goodness of fit and independence in two-way tables'. *Journal of the American Statistical Association*, vol.76, pp.221—30.

Raudenbush, SW & Bryk, AS (2001). *Hierarchical linear models: applications and data analysis methods* (vol. 1). Sage Publications.

Reinhard, S (1999). 'Econometric analysis of economic and environmental efficiency of Dutch dairy farms'. PhD thesis, Wageningen Universiteit.

Ross, J (2011). 'VET completion rates underestimated: skills council'. *Campus Review*, 28 March, vol.21, no.5.

Rumsey, D (2002). *Shaping the VET practitioner for the future*, Perth, Western Australian Department of Training.

Ryan, C (2002). Longer-term outcomes for individuals completing vocational education and training qualifications. Adelaide. NCVER.

—— (2011). *Year 12 completion and youth transitions. Longitudinal Surveys of Australian Youth. Research report 56*. Adelaide. NCVER.

Sadiq, S (2011). *The final frontier: a SAS approach to data envelopment analysis*. SAS paper 198—2011. SAS Global Forum.

Sales, AE, Plomondon, ME, Magid, DJ, Spertus, JA & Rumsfeld, JS (2004). 'Assessing response bias from missing quality of life data: The Heckman method'. *Health and Quality of Life Outcomes*, vol.2, no.1, p.49.

Sarle, WS (1983). *Cubic clustering criterion*. SAS technical report A—108. Cary, NC. SAS Institute Inc.

Sauvageot, C (2007). *Guidelines for developing indicators on technical and vocational education and training*. Torino. European Training Foundation.

Schrag, F (1992). 'In defense of positivist research paradigms', *Educational Researcher*, vol.21, no.5, pp.5—8.

Sizer, J, Spee, A & Bormans, R (1992). 'The role of performance indicators in higher education'. *Higher Education*, vol.24, no.2.

Smith, TMF (1983). 'On the validity of inferences from non-random sample'. *Journal of the Royal Statistical Society. Series A (General)*, pp.394—403.

Snell, D & Hart, A (2007). 'Vocational training in Australia: is there a link between attrition and quality?'. *Education + Training*, vol.49, no.6, pp.500—12.

Stanwick, J (2006). *Australian Qualifications Framework lower-level qualifications: outcomes for people over 25*. Adelaide. NCVER.

Stanwick, J, Hargreaves, J & Beddie, F (2009). *Assessing the impact of NCVER's research.* Adelaide. NCVER.

Stevens, PA (2005). 'A stochastic frontier analysis of English and Welsh universities'. *Education Economics*, vol.13, no.4, pp.355—74.

Taylor, B & Harris, G (2004). 'Relative efficiency among South African universities: a data envelopment analysis'. *Higher Education*, vol.47, no.1, pp.73—89.

Theodoridis, AM & Psychoudakis, A (2008). 'Efficiency measurement in Greek dairy farms: stochastic frontier vs. data envelopment analysis'. *International Journal of Economic Sciences and Applied Research*, no.2, pp.53—66.

Transparency Agenda in Vocational Education (2012). Retrieved 12 February 2013 from <http://www.dpmc.gov.au/publications/skills_for_all_australians/chapter6_greater_transparency_of_the_vet_sector.html>.

VFH (VET FEE HELP) (2015). Policy information. Retrieved 12 February 2015 from <http://studyassist.gov.au/sites/studyassist/helppayingmyfees/vet-fee-help/pages/vet-fee-help>.

Wadud, A & White, B (2000). 'Farm household efficiency in Bangladesh: a comparison of stochastic frontier and DEA methods'. *Applied Economics*, vol.32, no.13, pp.1665—73.

Waldman, DM (1982). 'A stationary point for the stochastic frontier likelihood'. *Journal of Econometrics*, vol.18, no.2, pp.275—9.

Wang, HJ & Schmidt, P (2002). 'One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels'. *Journal of Productivity Analysis*, vol.18, no.2, pp.129—44.

Ward, J (2008). *Measuring student satisfaction with vocational education and training services … and getting it right!*. Occasional paper. Adelaide. NCVER.

Whiteman, JL (1999). *The measurement of efficiency where there are multiple outputs* (no.g—134). Monash University, Centre of Policy Studies/ IMPACT Centre.

Winship, C & Radbill, L (1994). 'Sampling weights and regression analysis'. *Sociological Methods & Research*, vol.23, no.2, pp.230—57.

Woelert, P & Yates, L (2014). 'Too little and too much trust: performance measurement in Australian higher education'. *Critical studies in education*, (ahead-of-print), pp.1—15.

Worthington, AC (2001). 'An empirical survey of frontier efficiency measurement techniques in education'. *Education Economics*, vol.9, no.3, pp.245—68.

Worthington, AC & Lee, BL (2008). 'Efficiency, technology and productivity change in Australian universities, 1998—2003'. *Economics of Education Review*, vol.27, no.3, pp.285—98.

Wright, BD & Masters, GN (1982). *Rating scale analysis*. Chicago. MESA Press.

Wu, Y (1995). 'The productive efficiency of Chinese iron and steel firms: a stochastic frontier analysis'. *Resources Policy*, vol.21, no.3, pp.215—22.

Yorke, MR, Dove, A, Haslam, L, Hughes Jones, H, Longden, B, O'Connell, C, Typuszak, R & Ward, J (1997). *Undergraduate non-completion in England* (extended final report of a research project commissioned by HEFCE). Bristol. HEFCE.

Yorke, M (2000). 'The quality of the student experience: what can institutions learn from data relating to non-completion?'. *Quality in Higher Education*, vol.6, no.1, pp.61—75.

Zhang, Y & Bartels, R (1998). 'The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand'. *Journal of Productivity Analysis*, vol.9, no.3, pp.187—204.

7 Appendix

Tables to accompany Portfolio paper 1

Table A1 SOS 2011 raw data

	Graduates		Module completers		Graduates + module completers	
Institute	Respondents	Population	Respondents	Population	Respondents	Population
1	1,046	15,090	425	10,157	1,471	25,247
2	646	8,810	207	5,428	853	14,238
3	364	4,851	133	3,642	497	8,493
4	805	10,626	315	7,422	1,120	18,048
5	833	10,811	369	6,935	1,202	17,746
6	483	4,614	183	3,246	666	7,860
7	437	6,548	200	6,955	637	13,503
8	1,286	17,355	373	8,930	1,659	26,285
10	1,618	20,341	475	9,341	2,093	29,682
11	528	7,684	306	8,807	834	16,491
12	972	13,184	468	9,214	1,440	22,398
13	1,273	11,105	687	9,181	1,960	20,286
14	475	3,586	154	1,952	629	5,538
15	877	7,062	216	2,854	1,093	9,915
16	553	5,444	283	3,501	836	8,946
17	416	4,316	172	2,742	588	7,058
18	818	5,948	463	5,611	1,281	11,558
19	638	6,356	153	2,200	791	8,556
20	1,351	12,216	338	5,236	1,689	17,452
22	866	8,228	287	5,082	1,153	13,311
23	862	8,392	246	3,301	1,108	11,693
24	837	6,000	56	896	893	6,896
25	440	3,406	227	2,656	667	6,061
26	360	2,527	67	693	427	3,220
27	1,430	15,573	213	2,999	1,643	18,572
28	488	4,581	85	1,023	573	5,604
29	1,103	10,081	128	1,895	1,231	11,976
30	439	5,055	233	5,131	672	10,186
31	607	6,009	209	2,694	816	8,703
32	412	3,754	114	1,375	526	5,130
33	453	4,175	111	1,610	564	5,786
34	1,278	10,502	493	5,293	1,771	15,795
35	752	7,726	166	2,134	918	9,860
36	897	7,348	149	1,810	1,046	9,158
37	660	5,457	159	1,919	819	7,376
38	941	7,196	153	1,585	1,094	8,781
40	189	580	27	173	216	753
43	682	6,597	325	5,413	1,007	12,010
44	759	6,306	404	5,100	1,163	11,407
45	581	5,559	171	2,008	752	7,567
46	571	4,449	117	1,172	688	5,621
47	749	7,032	145	1,842	894	8,874

	Graduates		Module completers		Graduates + module completers	
Institute	Respondents	Population	Respondents	Population	Respondents	Population
48	1,614	10,713	315	2,541	1,929	13,254
49	534	3,305	193	1,906	727	5,211
50	1,227	8,519	302	2,450	1,529	10,968
51	2,112	10,543	404	2,924	2,516	13,467
52	435	2,896	95	698	530	3,593
53	467	2,674	144	944	611	3,619
55	542	3,243	145	888	687	4,131
56	233	1,381	109	786	342	2,167
57	314	1,757	82	581	396	2,338
58	449	3,108	175	1,448	624	4,556
60	261	1,505	83	594	344	2,099
61	76	194	7	16	83	210
64	985	7,428	593	7,057	1,578	14,484
65	903	7,496	500	5,846	1,403	13,342
66	1,048	6,126	747	8,023	1,795	14,149
70	1,404	11,680	300	2,846	1,704	14,525
71	1,038	5,428	191	1,573	1,229	7,002
74	55	743	26	230	81	973
75	1,255	5,233	349	1,775	1,604	7,008
77	1,176	5,622	372	3,063	1,548	8,685
78	2	9	124	780	126	790
109	33	93	1	7	34	100
110	299	2,374	52	836	351	3,210
Total	**48,237**	**424,550**	**15,514**	**218,968**	**63,751**	**643,518**

Table A2 Difference in salary estimates: adjusted v non-adjusted for selection bias

		Heckman		No Heckman				
Effect	Cate-gory	Estimate	Std error	Estimate	Std error	Difference estimate	t	Pr > t
Intercept	_	48361	3642	50163	3462	1802	0.36	0.719
IMRsalary	_	-5177	599	.	.	.	.	.
Age	_	345	12	42	3	-303	-25.22	<0.001
Sex (female)	0	-5480	316	-6933	320	-1453	-3.23	0.001
Sex (male)	1	0	.	0	.	0	.	.
Graduate	1	-1326	451	-714	440	612	0.97	0.332
Module Completer	2	0	.	0	.	0	.	.
Employed before	1	2822	2778	2364	2849	-458	-0.12	0.904
Employed before	2	-5498	2825	-7803	2897	-2305	-0.57	0.569
Employed before	3	-3777	2836	-6911	2908	-3134	-0.77	0.441
Employed before	4	-4012	3938	-5149	4039	-1137	-0.2	0.841
Employed before	999	0	.	0	.	0	.	.
Field of education	1	-4135	2011	-2170	2060	1964	0.68	0.497
Field of education	2	68	936	2111	938	2042	1.54	0.124
Field of education	3	892	665	2895	648	2003	2.16	0.031
Field of education	4	-1253	770	318	768	1571	1.44	0.15
Field of education	5	-167	812	2430	806	2597	2.27	0.023
Field of education	6	-937	734	1307	727	2244	2.17	0.03
Field of education	7	463	745	2884	738	2421	2.31	0.021
Field of education	8	-1164	675	818	658	1982	2.1	0.036
Field of education	9	-1912	724	-623	734	1288	1.25	0.211
Field of education	10	-2547	1076	-3935	1100	-1389	-0.9	0.368
Field of education	11	485	781	954	795	469	0.42	0.674
Field of education	12	0	.	0	.	0	.	.
Qualification level	1	255	752	-463	770	-718	-0.67	0.503
Qualification level	2	-362	725	-644	743	-282	-0.27	0.787
Qualification level	3	-2624	716	-5255	724	-2631	-2.58	0.01
Qualification level	4	-5767	769	-8784	761	-3017	-2.79	0.005
Qualification level	5	-1802	1115	-5311	1088	-3509	-2.25	0.024
Qualification level	6	966	654	1322	670	356	0.38	0.704
Qualification level	7	0	.	0	.	0	.	.
Occupation	10	10266	2258	14395	2312	4130	1.28	0.201
Occupation	11	11425	1653	15556	1691	4131	1.75	0.08
Occupation	12	-8446	1756	-4565	1794	3881	1.55	0.121
Occupation	13	12957	1171	16102	1196	3145	1.88	0.06
Occupation	14	-1632	1286	1049	1315	2681	1.46	0.144
Occupation	20	6138	3481	8618	3569	2480	0.5	0.617
Occupation	21	-5268	2465	-2729	2527	2540	0.72	0.472
Occupation	22	6073	1229	8276	1257	2203	1.25	0.211
Occupation	23	6188	1213	7743	1241	1555	0.9	0.368
Occupation	24	8200	1251	10715	1280	2515	1.41	0.159
Occupation	25	11608	1350	14295	1382	2687	1.39	0.165
Occupation	26	6301	1548	7584	1585	1283	0.58	0.562
Occupation	27	3449	1510	6138	1544	2689	1.24	0.215
Occupation	30	1471	2114	3438	2167	1967	0.65	0.516
Occupation	31	-1471	1184	-188	1213	1282	0.76	0.447
Occupation	32	-2921	1125	-2817	1153	103	0.06	0.952

		Heckman		No Heckman				
Effect	Cate-gory	Estimate	Std error	Estimate	Std error	Difference estimate	t	Pr > t
Occupation	33	-6568	1178	-6391	1207	177	0.11	0.912
Occupation	34	-695	1124	-39	1152	656	0.41	0.682
Occupation	35	-7745	1479	-7458	1514	286	0.14	0.889
Occupation	36	-10967	1427	-11001	1462	-33	-0.02	0.984
Occupation	39	-3575	1311	-3133	1344	442	0.24	0.81
Occupation	40	-4779	5313	-2057	5449	2722	0.36	0.719
Occupation	41	-4217	1315	-2290	1343	1926	1.03	0.303
Occupation	42	-9123	1256	-7597	1285	1526	0.85	0.395
Occupation	43	-9305	1814	-9200	1860	105	0.04	0.968
Occupation	44	2217	1545	3511	1583	1294	0.59	0.555
Occupation	45	-8743	1622	-8308	1662	435	0.19	0.849
Occupation	50	-3354	2017	-1613	2068	1741	0.6	0.549
Occupation	51	4670	1263	7322	1292	2652	1.47	0.142
Occupation	52	-4403	1629	-2757	1670	1647	0.71	0.478
Occupation	53	-8813	1221	-7678	1250	1136	0.65	0.516
Occupation	54	-9331	1384	-9568	1417	-238	-0.12	0.904
Occupation	55	-6054	1340	-3977	1370	2078	1.08	0.28
Occupation	56	-10264	1985	-8569	2034	1694	0.6	0.549
Occupation	59	-4810	1297	-2901	1328	1909	1.03	0.303
Occupation	60	-8556	5522	-5836	5664	2720	0.34	0.734
Occupation	61	-5944	1790	-4621	1830	1322	0.52	0.603
Occupation	62	-10575	1392	-10142	1425	433	0.22	0.826
Occupation	63	-9336	2028	-9540	2078	-203	-0.07	0.944
Occupation	70	-6076	2951	-4097	3025	1979	0.47	0.638
Occupation	71	-782	1362	1062	1393	1844	0.95	0.342
Occupation	72	-5320	1520	-3434	1555	1886	0.87	0.384
Occupation	73	-9149	1606	-5967	1642	3182	1.39	0.165
Occupation	74	-11698	1722	-10045	1764	1653	0.67	0.503
Occupation	80	-9720	3217	-8917	3262	803	0.18	0.857
Occupation	81	-11084	1819	-7869	1861	3215	1.24	0.215
Occupation	82	-6660	1401	-5366	1434	1293	0.65	0.516
Occupation	83	-8476	1452	-6950	1487	1526	0.73	0.465
Occupation	84	-11445	1461	-11049	1497	396	0.19	0.849
Occupation	85	-13787	1983	-11599	2031	2188	0.77	0.441
Occupation	89	-9284	1487	-7718	1523	1566	0.74	0.459
Occupation	999	0	.	0	.	0	.	.
Industry	_	5811	2194	5046	2251	-764	-0.24	0.81
Industry	_	10697	4968	8028	5095	-2669	-0.38	0.704
Industry	3	13488	2969	12802	3033	-686	-0.16	0.873
Industry	4	17124	3385	17796	3472	672	0.14	0.889
Industry	5	8359	2547	9722	2612	1364	0.37	0.711
Industry	6	34648	2769	35712	2832	1064	0.27	0.787
Industry	7	29615	2291	31215	2349	1600	0.49	0.624
Industry	8	30135	2028	30069	2080	-65	-0.02	0.984
Industry	9	16200	3598	17041	3691	842	0.16	0.873
Industry	10	31796	3473	31345	3563	-451	-0.09	0.928
Industry	11	7561	2119	7848	2173	287	0.09	0.928
Industry	12	6361	3082	6389	3161	28	0.01	0.992

		Heckman		No Heckman				
Effect	Cate-gory	Estimate	Std error	Estimate	Std error	Difference estimate	t	Pr > t
Industry	13	-5265	3002	-5108	3079	157	0.04	0.968
Industry	14	-1406	2561	-2160	2627	-753	-0.21	0.834
Industry	15	20597	3633	20659	3727	62	0.01	0.992
Industry	16	1268	2779	1853	2850	585	0.15	0.881
Industry	17	22941	3624	23376	3717	435	0.08	0.936
Industry	18	14073	2615	14424	2682	350	0.09	0.928
Industry	19	8044	2936	8860	3012	816	0.19	0.849
Industry	20	7589	2579	7967	2646	378	0.1	0.92
Industry	21	15457	2266	16124	2324	667	0.21	0.834
Industry	22	3638	2185	2884	2241	-753	-0.24	0.81
Industry	23	9656	2186	10354	2243	698	0.22	0.826
Industry	24	7919	2117	7980	2171	61	0.02	0.984
Industry	25	-5091	2554	-6551	2619	-1460	-0.4	0.689
Industry	26	20488	2134	21254	2188	766	0.25	0.803
Industry	27	28721	3663	30337	3757	1616	0.31	0.757
Industry	28	10911	2262	12183	2320	1271	0.39	0.697
Industry	29	11934	3114	12870	3194	936	0.21	0.834
Industry	30	12360	2042	11982	2094	-378	-0.13	0.897
Industry	31	14584	2099	15057	2153	473	0.16	0.873
Industry	32	8582	1958	7612	2008	-970	-0.35	0.726
Industry	33	5476	2713	5228	2783	-248	-0.06	0.952
Industry	34	4653	2742	4037	2813	-617	-0.16	0.873
Industry	35	7676	4945	7451	5072	-225	-0.03	0.976
Industry	36	4812	3165	3474	3246	-1338	-0.3	0.764
Industry	37	4140	3028	4034	3106	-106	-0.02	0.984
Industry	38	7709	3787	6248	3884	-1461	-0.27	0.787
Industry	39	-929	2490	-1673	2554	-744	-0.21	0.834
Industry	40	1157	4042	781	4147	-376	-0.06	0.952
Industry	41	694	2309	-354	2367	-1048	-0.32	0.749
Industry	42	1259	2050	423	2103	-836	-0.28	0.779
Industry	43	-3307	4819	-4427	4943	-1120	-0.16	0.873
Industry	44	2428	2313	3251	2373	824	0.25	0.803
Industry	45	2064	2111	906	2165	-1158	-0.38	0.704
Industry	46	9948	2197	10350	2253	402	0.13	0.897
Industry	47	19554	2721	21268	2791	1714	0.44	0.66
Industry	48	20616	2812	21036	2876	419	0.1	0.92
Industry	49	15751	3035	15567	3113	-183	-0.04	0.968
Industry	50	3651	3321	4013	3407	362	0.08	0.936
Industry	51	6540	2912	10058	2985	3517	0.84	0.401
Industry	52	20250	2541	20850	2606	600	0.16	0.873
Industry	53	9158	2588	8901	2654	-256	-0.07	0.944
Industry	54	4822	3668	4710	3762	-111	-0.02	0.984
Industry	55	6593	4092	3423	4196	-3170	-0.54	0.589
Industry	56	8069	4089	9148	4194	1079	0.18	0.857
Industry	57	21299	13405	19085	13750	-2214	-0.12	0.904
Industry	58	11938	2413	12380	2475	441	0.13	0.897
Industry	59	6958	4609	5693	4728	-1265	-0.19	0.849
Industry	60	1154	3135	3339	3215	2184	0.49	0.624

		Heckman		No Heckman				
Effect	Cate-gory	Estimate	Std error	Estimate	Std error	Difference estimate	t	Pr > t
Industry	62	12236	2265	11667	2323	-569	-0.18	0.857
Industry	63	11106	2689	9885	2758	-1221	-0.32	0.749
Industry	64	7167	3303	6214	3388	-952	-0.2	0.841
Industry	66	5905	3095	4369	3159	-1536	-0.35	0.726
Industry	67	3803	2349	3068	2408	-735	-0.22	0.826
Industry	69	6988	1979	6340	2030	-648	-0.23	0.818
Industry	70	7123	2264	5432	2321	-1691	-0.52	0.603
Industry	72	3776	2109	3051	2163	-725	-0.24	0.81
Industry	73	5248	2186	5907	2242	659	0.21	0.834
Industry	75	12218	1966	13152	2016	934	0.33	0.741
Industry	76	14975	2259	14693	2316	-282	-0.09	0.928
Industry	77	14526	2136	15622	2190	1096	0.36	0.719
Industry	80	9979	1992	12156	2042	2177	0.76	0.447
Industry	81	7610	2020	10208	2071	2599	0.9	0.368
Industry	82	2992	2522	5048	2587	2056	0.57	0.569
Industry	84	6449	2105	7990	2159	1541	0.51	0.61
Industry	85	8114	2045	8624	2097	510	0.17	0.865
Industry	86	1631	2101	4310	2154	2679	0.89	0.373
Industry	87	2885	2002	3906	2053	1021	0.36	0.719
Industry	89	3986	2380	4495	2441	509	0.15	0.881
Industry	90	4218	3519	5064	3610	846	0.17	0.865
Industry	91	5123	2427	3976	2489	-1148	-0.33	0.741
Industry	92	10232	3268	10621	3352	389	0.08	0.936
Industry	94	1583	2018	644	2069	-938	-0.32	0.749
Industry	95	-3924	2162	-4952	2217	-1029	-0.33	0.741
Industry	96	-9577	5786	-12991	5934	-3414	-0.41	0.682
Industry	999	0	.	0	.	0	.	.
Income deviation	_	0	0	0	0	0	0.07	0.944

Notes: Highlighted rows = Significant difference between Heckman and non-Heckman model.

Occupation and Industry classification as per Australia New Zealand Standard Industry Classification (ANZSIC) 2006.

Field of Education:
1 = 'Natural and Physical Sciences'
2 = 'Information Technology'
3 = 'Engineering and Related Technologies'
4 = 'Architecture and Building'
5 = 'Agriculture, Environmental and Related Studies'
6 = 'Health'
7 = 'Education'
8 = 'Management and Commerce'
9 = 'Society and Culture'
10 = 'Creative Arts'
11 = 'Food, Hospitality and Personal Services'
12 = 'Mixed Field Programmes'

Employed before Training:
1 = 'Wage or salary earner'
2 = 'Conducting own business - with employees'
3 = 'Conducting own business - without employees'
4 = 'Helper not receiving wages'

Qualification level:
1 = 'Bachelor Degree or higher'
2 = 'Advanced Diploma or Associate Degree'
3 = 'Diploma or Associate Diploma'
4 = 'Certificate IV (or Advanced Certificate/Technician)'
5 = 'Certificate III (or Trade Certificate eg. apprenticeship)'
6 = 'Certificate II'
7 = 'Certificate I'

Generally: 999 = 'Not Known or Not Stated' ;