

Modelling the reasons for training choices: Technical paper

ANDREW SMITH, EDDIE OCZKOWSKI & MARK HILL

CENTRE FOR ORGANISATIONAL PERFORMANCE,
ETHICS AND LEADERSHIP, CHARLES STURT
UNIVERSITY

This document was produced by the author(s) based on their research for the report *Reasons for training: Why Australian employers train their workers*, and is an added resource for further information. The report is available on NCVER's website: [<http://www.ncver.edu.au>](http://www.ncver.edu.au)

The views and opinions expressed in this document are those of the author(s) and do not necessarily reflect the views of the Australian Government, state and territory governments or NCVER. Any errors and omissions are the responsibility of the author(s).

© Australian Government, 2009

This work has been produced by the National Centre for Vocational Education Research (NCVER) on behalf of the Australian Government and state and territory governments with funding provided through the Australian Department of Education, Science and Training. Apart from any use permitted under the *Copyright Act 1968*, no part of this publication may be reproduced by any process without written permission. Requests should be made to NCVER.

Modelling the Reasons for Training Choices

Introduction

This report provides the technical details on the modelling aspects of identifying significant drivers for the reasons for using certain types of training and for the choice of training types. The employed data is from the 2005 Survey of Employer Use and Views of the VET system (SEUV). The data has previously been analysed in NCVER (2006).

This report focuses on two broad research questions. First, what are the drivers of the reasons (or what explains the reasons) employers cite for using particular training types. Second, what explains the observed choice employers make in choosing between nationally recognised training and unaccredited training. These two research questions are analysed in turn in sections three and four.

Data Issues

The SEUV sample data consists of observations from 4601 interviews. The survey seeks to be representative of a larger group of 892,275 employers as defined by the Australian Business Register. Weights have been constructed by the ABS and described by Adena (2006) to ensure comparability between the sample data and the population of all employers according to location (state), industry type and employee size. In order to identify the statistically important drivers for the various models and still maintain the weighting of the data (to assure some concordance with the underlying population of Australian businesses) all analysed data are weighted and re-scaled. The re-scaling of the weights is designed to produce a sample size consistent with the sample collected but still maintaining the relative weighting characteristics. Re-scaling is performed using: $\text{weight} * (\text{sample size} / \text{population size})$.

For missing values, listwise deletion (using cases only where all observations exist for all the variables in the proposed model) is used to develop the various regression models. As a consequence the number of cases for each model differs and is less than the overall sample size. Initially convergence problems when estimating the models existed for the 'industry type' and 'organisational status' variables. The problem resulted from some cells being too small. As a consequence observations for the 'other' category for industry type and organisational status were deleted. This resulted in the deletion of only nine cases, resulting in an available sample size of 4592.

Modelling the reasons cited for using training types

Modelling Approach

The SEUV data set provides information on the reasons why employers have used certain training types. These reasons effectively indicate what they [expect to] get from a specific training type. We will model the drivers for these reasons for choosing a specific training type.

Four sets of models will be developed, one for each of the following dependent variables:

- ✧ Reasons for vocational qualifications as a job requirement
- ✧ Reasons for using apprenticeships/traineeships
- ✧ Reasons for using nationally recognised training
- ✧ Reasons for using unaccredited training

The variables to be investigated as potentially important drivers for the reasons for training choice are categorised into three groups.

Organisational Characteristics

- ✧ Industry type: 17 different ANZSIC divisions.
- ✧ Organisational status: four types (private for profit, private not for profit, government business enterprise, other government enterprise).
- ✧ Organisational employee size (small 1-9, medium 10-99, large 100+).
- ✧ Permanent/ongoing employees: three categories,
 - ◆ low 0-25%, in the sample this consists of approximately 25% of organisations,
 - ◆ medium 26%-99%, in the sample this consists of approximately 50% of organisations,
 - ◆ high 100%, in sample this consists of approximately 25% of organisations.
- ✧ Occupational emphasis: three mutually exclusive but non-exhaustive classifications,
 - ◆ blue collar (an organisation with more than 50% of machinery operators and drivers, labourers) in the sample this consists of approximately 20% of organisations,
 - ◆ white collar (an organisation with more than 50% of clerical and administrative staff) in the sample this consists of approximately 10% of organisations,
 - ◆ knowledge workers (an organisation with more than 50% of professionals, technicians and tradespersons) in the sample this consists of approximately 27% of organisations.
- ✧ RTO status, whether the organisation is a registered training organisation.

Strategic

- ✧ Existence of a business plan.
- ✧ Staff training as part of a business plan.
- ✧ Importance of training (five point scale: 1 very unimportant, to 5 very important).

Skill Recruitment

- ✧ Current skill level (three point scale: 1 below what is required, to 3 above what is required).
- ✧ Recruitment difficulties (three point scale: 1 no difficulty, to 3 a lot of difficulty).

Data Analysis Techniques

Numerous reasons are listed for each of the training choices and many employers list more than one reason for a training choice. Table 1 lists information on the number of cited choices and their frequency for each of the four analysed training types. In general, approximately 60% of employers cite a single reason, approximately 30% cite two reasons and approximately 10% of employers cite three or more reasons for choosing a training type. This data prohibits the mutually exclusive characterisation of employers based on a single (or most important) reason. As a consequence cluster analysis is employed to group employers according to reasons chosen.

Table 1 **Number and Frequency of cited Reasons for using Training**

Training Type	Total Number of Reasons	Percentage of Organisations Citing Number of Reasons		
		One Reason	Two Reasons	Three or More Reasons
Vocational Qualifications	7	67.0	23.1	9.1
Apprenticeships/Traineeships	15	56.2	30.1	13.7
Nationally Recognised Training	16	64.5	25.1	10.4
Unaccredited Training	12	57.2	30.4	12.4

Cluster Analysis

The role of cluster analysis is to group cases of a similar nature into distinct exhaustive but mutually exclusive categories (Hair et al. 2006). In modelling the reasons for training we need to be able to categorise the employers into unique non-overlapping categories even though on average 40% of employers cite more than one reason for using a training type. The survey did not ask employers to rank the cited reasons in order of importance and hence this and other types of information cannot be employed to uniquely classify the reasons. To this extent the cluster analysis by definition will produce some clusters which have employers who cite two or more reasons for choosing a training type.

Initially we subjected all the individual cited reasons for training choices to cluster analysis, however, this produced clusters which were too complex to understand given the large number of cited reasons for each training choice, see table 1. To simplify analysis and understanding, we combined reasons a priori of similar meaning before undertaking further analysis. The combined reasons and their meanings are provided in table 2. These variables are used in the cluster analysis.

Table 2 Combined Reasons for use Training Choice Analysis

Training Type	Combined No. of Reasons	Combined Reasons
Vocational Qualifications	4	Skills: Provides skills required for the job. Standards: Maintain professional standards/ Meeting industry standards. Regulations: Legislative, regulatory or licensing requirements. Award or enterprise bargaining agreements. Competition: Improving quality of services and goods provided. Responding to new technology. To remain competitive.
Apprenticeships/ Traineeships	6	Specific Skills: Gain skills specific to business/Train to own requirements Skilling Staff: To get skilled staff / improve staff skills. Staff gain a nationally recognised qualification. Specific Role: To fill a specific role/ Need more staff. Government /Regulatory requirements. Ethical: Give young people a head start. Support our industry/ Give something back to our industry. Improving indigenous skills/ Employing indigenous people. Cost: Financial incentives. Cheap labour/ cost effective. To do some of the simpler tasks. Practice & Culture: Part of business 'training culture'/The way we do things. Usual business practice/ Have always employed apprentices. Work force planning. Succession. Ageing workforce. To help retain staff.
Nationally Recognised Training	5	Regulations: Legislative, regulatory or licensing requirements. Award or enterprise bargaining agreements. Skills: Provides skills required for the job. Standards: Maintain professional standards/ Meeting industry standards. Competition: Improving quality of services/goods provided. Responding to new technology. To remain competitive. To receive a subsidy / Cost effective. Helps business growth /Adds value to business. Human Resources: Developing and maintaining a flexible and responsive workforce. Employee/staff requested. To improve staff morale / self esteem. To help employee retention. Career development / To increase or update skills. Formalise qualifications / skills. Allow them to move around the industry / around Australia
Unaccredited Training	5	Skills: Provides skills required for the job. To increase/update skills. Competition: Improving quality of services/goods provided Responding to new technology. To remain competitive. Cost effective. Standards: Maintain professional standards/ Meeting industry standards. Responsive Workforce: Developing and maintaining a flexible and responsive workforce. To meet highly specific training needs. Regulations: Legislative, regulatory or licensing requirements. Award or enterprise bargaining agreements. There is no accredited training for this industry.

We employ the two-step cluster analysis procedure available in SPSS (ver14). The procedure initially pre-clusters the data using a cluster feature tree into many small sub-clusters. Using these sub-clusters the second step determines the final clusters using an agglomerative hierarchical method. The technique is particularly efficient for large data sets and can handle categorical data such as our cited reasons for training by using the log-likelihood as a distance measure. Simulation performance reported in SPSS (2001) and Chiu et al. (2001) suggests the procedure may perform accurately under a variety of conditions.

Initially, we employed Schwarz's Bayesian criterion to optimally choose the number of clusters. This resulted in an optimum number of clusters of ten for vocational qualifications, six for apprenticeships/traineeships and five clusters for both unaccredited training and nationally recognised training. A large number of clusters are not particularly useful in our context as it makes the interpretation of the modelling too complex. To maintain consistency across training choices and to get an appropriate balance between complexity and richness we choose five clusters for each training type for modelling purposes.

Logit Models

The dependent variables based on the cluster analysis which relate to the cited reasons for using a certain training type are categorical variables without any natural order. As consequence the multi-nominal logit model (MNL) is the natural tool of analysis to develop an appropriate model and identify any important drivers for the reasons for training within each of the four broad training types. However, the MNL provides the identification of drivers for all possible pairwise comparisons, in our case with five clusters this implies ten comparisons. Once again this complicates description and makes interpretation of results difficult. As an alternative we employ binary logit models for each training choice. With five clusters this involves five binary logits. Effectively each logit is a comparison of the chosen cluster against all other clusters combined. In essence, binary logits average the results of the MNL.

To better identify the individual significant drivers of reasons for training choices and avoid the consequences of multi-collinearity, rather than present estimates for all the potential drivers (section 3.1) we present significant results based on stepwise procedures. To assure some sense of robustness in the modelling approach, we employed both forward and backward stepwise procedures using the likelihood ratio criteria with probabilities of 0.05 for entry and 0.10 for removal. For nine of the 20 estimated models, the forward and backward procedures identified the same models. In the other cases the backward procedure identified one or more additional drivers. To avoid the omitted variable bias problem and to use more rather than less information, our subsequent results are based on the backward stepwise procedure.

The B coefficients (the log-odds), presented in tables 4, 6, 8 and 10, are the log-odds of the listed cluster compared to all other clusters combined. These log-odds can be converted into odds by the transformation ($\exp(B)$), see for example Long (1997). In addition to the estimated log-odds and their t-ratios (coefficients / standard errors), we present goodness of fit measures (proportion of correct predictions and the Nagelkerke R^2 measure) and statistical tests for continuous variables and categorical variables. The interpretation of the business plan variables requires some comment. If both the existence a business plan and staff training as part of the business plan variables are retained, then it is the sum of the two coefficients which measures the effect of staff training as part of the business plan, as staff training can only be part of a plan if a plan exists. If only one of the two business plan variables is retained then its coefficient measures the effect of that variable alone.

Reasons for using Vocational Qualifications as a Job Requirement

The results of the cluster analysis for vocational qualifications as a job requirement are presented in table 3. For all cluster analysis results, the reasons are listed in order of frequency and the clusters in order of simplicity of structure.

Table 3 Cluster Analysis: Reasons for Vocational Qualifications as a job requirement

Reasons	Cluster					Total
	1	2	3	4	5	
Skills	451 (100)			309 (100)	125 (63)	885
Standards			400 (100)	236 (76)	88 (44)	724
Regulations		402 (100)	102 (26)	141 (46)	39 (20)	684
Competition					198 (100)	198
Number of Employers	451	402	400	309	198	

Note. Numbers in parentheses represent the percent of employers in the cluster who cite the stated reason. Entries are **bolded** if cited by 50% or more of employers.

We provide the following interpretation of the identified clusters.

Cluster 1 Skills: Skills required only.

Cluster 2 Regulations: Regulation related reasons only.

Cluster 3 Standards and Regulations: All cite standards and about quarter regulations.

Cluster 4 Skills, Standards and Regulations: All cite skills, three quarters standards and about half regulations.

Cluster 5 Competition, Skills and other reasons: All cite competition and about two thirds skills, the other two reasons are also cited to some but minor extent.

Note, cluster five might be seen as a ‘residual’ cluster with all the reasons cited appearing. An attempt was made to relocate those employers into other single reason clusters but without any success. In essence, moving employers from cluster five to the other clusters necessarily carries across some of the other reasons and therefore the single reason clusters no longer maintained their simplicity of structure. Even so, the residual cluster can be labelled in terms of the dominant reasons, in this case competition and skills. This comment about the ‘residual’ cluster also applies to the other training choices.

The binary logit estimates are presented in table 4. The models are based on 1760 cases, with the proportion of correct predictions ranging from 0.75 to 0.89 and pseudo R^2 s ranging from 0.11 to 0.21. Industry type, employee size and importance of training are significant drivers for all clusters. RTO status and recruitment difficulties are unimportant drivers for all training choice reasons.

Table 4 Logit Parameter Estimates: Reasons for Vocational Qualifications as a Job Requirement

Variable	Cluster 1 Skills		Cluster 2 Regulations		Cluster 3 Standards & Regulations		Cluster 4 Skills, Standards & Regulations		Cluster 5 Competition, Skills & Other Reasons	
	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio
Constant	-2.361*	-5.31	-3.745*	-5.01	-0.753	-0.93	0.525	0.61	-1.583*	-2.09
Industry Type (Control: Professional & Other Services)	0.000*		0.000*		0.000*		0.000*		0.000*	
Agriculture, Forestry and Fishing	-0.328	-0.96	0.609	0.69	-0.819*	-2.10	1.514*	3.73	-1.106*	-2.22
Mining	0.521	0.72	-0.321	-0.27	0.487	0.67	-0.523	-0.38	-1.755	-1.28
Manufacturing	-0.186	-0.66	-1.105*	-2.32	-0.604*	-2.03	1.381*	3.97	-0.002	-0.01
Electricity, Gas and Water Supply	-3.351*	-2.84	1.668*	3.84	0.574	1.51	-1.338	-1.86	0.122	0.25
Construction	0.274	1.00	0.640	1.69	-1.084*	-3.42	0.183	0.50	0.175	0.49
Wholesale Trade	1.663*	4.33	-1.359	-1.84	-1.300*	-2.54	0.556	0.96	-1.607*	-2.48
Retail Trade	-0.105	-0.38	0.406	1.08	-0.495	-1.67	1.174*	3.38	-1.209*	-3.02
Accommodation, Cafes and Restaurants	-0.230	-0.61	0.830	1.90	-0.478	-1.24	0.858	1.72	-1.141*	-2.41
Transport and Storage	-1.055*	-2.37	0.563	1.20	0.287	0.78	0.547	1.15	-0.408	-0.82
Communication Services	-1.037	-1.73	0.364	0.59	-1.065	-1.78	1.850*	3.71	-0.582	-0.88
Finance and Insurance	-0.066	-0.22	1.554*	4.03	-0.604	-1.82	-0.682	-1.42	-0.964*	-2.10
Property and Business Services	-0.479	-1.60	0.991*	2.64	-0.031	-0.11	0.430	1.15	-1.194*	-2.76
Government Administration and Defence	-0.342	-0.41	0.380	0.38	-0.583	-0.49	1.285	1.30	-1.582	-1.38
Education	-0.610	-1.23	2.214*	4.55	-0.432	-0.81	-1.478*	-2.25	-1.368*	-2.23
Health and Community Services	-0.390	-1.33	1.636*	4.43	-0.662*	-2.08	0.170	0.44	-2.067*	-4.41
Cultural and Recreational Services	-1.141*	-2.59	0.891*	2.09	0.394	1.11	-0.743	-1.22	-0.204	-0.48
Organisational Status (Control: Other Gov't Org'n)			0.021*		0.004*		0.000*			
Private 'for Profit'			-0.412	-0.91	0.607	1.00	-0.741	-1.27		
Private 'Not for Profit'			-0.826	-1.79	-0.426	-0.68	1.081	1.84		
Gov't Business Enterprise			1.220	1.48	-0.304	-0.28	-1.595	-1.25		
Employee Size (Control: large)	0.010		0.005*		0.054		0.099		0.000*	
Small	0.388	1.35	1.298*	3.24	-0.475	-1.63	-0.137	-0.41	-1.024*	-3.16
Medium	-0.002	-0.01	1.255*	3.11	-0.182	-0.63	-0.466	-1.38	-0.431	-1.37
Permanency of Employees (Control: High)					0.083				0.019*	
Low					0.211	0.68			0.543	1.40
Medium					0.304*	2.22			0.485*	2.71
Occupational Type										
Blue Collar Organisation					-0.415	-1.67				
White Collar Organisation			-1.029*	-4.12	0.779*	3.49	1.072*	3.24	-0.785	-1.74
Knowledge Worker Organisation			-0.648*	-4.56			1.425*	8.33	-0.579*	-3.08
Strategic and Skill Variables										
Registered Training Organisation										
Existence of a Business Plan					1.090*	3.85	-1.959*	-3.79		
Staff Training as part of Business Plan					-0.839*	-3.42	1.444*	2.89	0.729*	2.96
Importance of Training	-0.319*	-4.49	0.255*	2.74	-0.290*	-3.54	0.358*	3.65	0.322*	2.56
Current Skill Level	-0.402*	-3.59			0.436*	3.79	-0.475*	-3.63	0.298*	2.00
Recruitment Difficulties										
Nagelkerke R ²	0.128		0.178		0.112		0.206		0.146	
Proportion of Correct Predictions	0.751		0.784		0.782		0.833		0.886	

Significance levels are presented for retained multiple group categorical variables next to controls. * denotes significance at a 5% level.

Reasons for using Apprenticeships/Traineeships

The results of the cluster analysis for apprenticeships/traineeships are presented in table 5.

Table 5 Cluster Analysis: Reasons for using Apprenticeships/Traineeships

Reasons	Cluster					Total
	1	2	3	4	5	
Specific Skills	290 (100)	81 (35)	103 (37)	67 (27)	127 (31)	668
Skilling Staff		232 (100)	48 (17)	55 (22)	115 (28)	450
Specific Role			280 (100)	58 (24)	107 (26)	445
Ethical				246 (100)	94 (23)	340
Cost					239 (58)	239
Practice & Culture					212 (52)	212
Number of Employers	290	232	280	246	411	

Note. Numbers in parentheses represent the percent of employers in the cluster who cite the stated reason. Entries are **bolded** if cited by 50% or more of employers.

We provide the following interpretation of the identified clusters.

- Cluster 1 Specific Skills:** Specific skills required only.
- Cluster 2 Skilling Staff:** All cite skilling staff and about one third specific skills which are related to skilling staff.
- Cluster 3 Specific Role and Skills:** All cite specific role, about one third specific skills and some skilling staff. The two less cited reasons are termed skills generically.
- Cluster 4 Ethical:** All cite moral related reasons and about a quarter cite three other reasons which generically relate to skills.
- Cluster 5 Cost, Practice & Culture and other reasons:** About half cite both cost and practice & culture, and the other four reasons are cited but to a lesser extent.

The binary logit estimates are presented in table 6. The models are based on 1459 cases, with the proportion of correct predictions ranging from 0.72 to 0.85 and pseudo R^2 s ranging from 0.11 to 0.24. Only industry type is a significant driver for all clusters. All potential drivers listed in section 3.1 are important for at least one combined reason.

Table 6 Logit Parameter Estimates: Reasons for using Apprenticeships/Traineeships

Variable	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Specific Skills		Skilling Staff		Specific Role & Skills		Moral & Skills		Cost, Practice & Culture & Other Reasons	
	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio
Constant	-3.326*	-2.60	-4.656*	-3.75	1.090	1.69	0.809	1.07	-2.124*	-5.34
Industry Type (Control: Professional & Other Services)	0.000*		0.000*		0.000*		0.000*		0.013*	
Agriculture, Forestry and Fishing	0.414	0.79	2.443*	2.17	-0.064	-0.14	-1.247*	-2.23	-0.479	-1.15
Mining	-2.009	-0.79	3.529*	2.52	-0.634	-0.54	-0.074	-0.06	-0.680	-0.60
Manufacturing	0.441	0.95	1.974	1.79	-0.006	-0.01	-0.293	-0.67	-0.820*	-2.19
Electricity, Gas and Water Supply	0.674	1.21	0.695	0.54	0.198	0.42	-0.279	-0.52	-0.244	-0.55
Construction	0.624	1.37	0.880	0.79	-1.398*	-3.23	0.588	1.44	-0.028	-0.08
Wholesale Trade	0.950	1.66	1.178	0.97	-0.044	0.08	0.216	0.39	-0.796	-1.44
Retail Trade	-0.221	-0.47	1.681	1.53	0.412	1.04	-0.166	-0.40	-0.360	-1.02
Accommodation, Cafes and Restaurants	0.559	1.06	1.424	1.25	-0.249	-0.48	-1.152	-1.95	0.347	0.81
Transport and Storage	-0.162	-0.26	1.404	1.19	0.216	0.44	-1.547*	-2.11	0.480	1.10
Communication Services	1.191	1.49	0.168	0.10	1.671*	2.45	-1.890	-1.47	-1.219	-1.39
Finance and Insurance	0.808	1.45	2.657*	2.39	-0.291	0.57	-2.617*	-2.88	-0.467	-1.08
Property and Business Services	1.954*	4.17	0.544	0.47	-0.946	-1.95	-1.451*	-2.70	-0.237	-0.62
Government Administration and Defence	1.039	0.77	1.425	0.97	-0.077	-0.08	-1.364	-1.35	-0.589	-0.69
Education	0.781	0.86	1.949	1.49	0.357	0.49	-1.254	-1.48	-1.214	-1.50
Health and Community Services	-0.196	-0.33	1.465	1.30	0.267	0.59	-0.917	-1.73	-0.083	-0.20
Cultural and Recreational Services	1.230*	2.08	0.522	0.42	0.272	0.50	-1.744*	-2.51	0.086	0.18
Organisational Status (Control: Other Gov't Organisation)	0.016*						0.000*			
Private 'for Profit'	1.143	1.06					-2.444*	-3.73		
Private 'Not for Profit'	-0.262	-0.23					-0.627	-0.99		
Gov't Business Enterprise	-0.944	-0.46					-1.901	-1.74		
Employee Size (Control: large)			0.006*							
Small			0.132	0.30						
Medium			0.684	1.59						
Permanency of Employees (Control: High)	0.004*									
Low	-0.066	-0.18								
Medium	0.525*	3.24								
Occupational Type										
Blue Collar Organisation			-0.659*	-2.34	0.774*	3.11				
White Collar Organisation	-1.362*	-3.28			0.881*	2.52				
Knowledge Worker Organisation			-1.065*	-5.14	0.790*	4.49			0.233	1.66
Strategic and Skill Variables										
Registered Training Organisation							-0.660*	-2.09	0.636*	3.01
Existence of a Business Plan									-0.997	-1.86
Staff Training as part of Business Plan	0.432*	2.44	-1.309*	-7.19					1.563*	3.01
Importance of Training							-0.173	-1.75		
Current Skill Level	0.618*	4.87	-0.560*	-4.27						
Recruitment Difficulties	0.216*	2.54	0.175	1.84	0.244*	2.94			-0.504*	-6.72
Nagelkerke R ²	0.163		0.242		0.107		0.144		0.113	
Proportion of Correct Predictions	0.816		0.846		0.806		0.834		0.720	

Significance levels are presented for retained multiple group categorical variables next to controls. * denotes significance at a 5% level.

Reasons for using Nationally Recognised Training

The results of the cluster analysis for nationally recognised training are presented in table 7.

Table 7 Cluster Analysis: Reasons for using Nationally Recognised Training

Reasons	Cluster					Total
	1	2	3	4	5	
Regulations	52 (25)				546 (100)	598
Skills			334 (100)	89 (36)	72 (13)	495
Standards	208 (100)	47 (33)	40 (12)	58 (23)	46 (8)	399
Competition		142 (100)	79 (24)	86 (35)	47 (9)	354
Human Resources				249 (100)	38 (7)	287
Number of Employers	208	142	334	249	546	

Note. Numbers in parentheses represent the percent of employers in the cluster who cite the stated reason. Entries are **bolded** if cited by 50% or more of employers.

We provide the following interpretation of the identified clusters.

- Cluster 1 Standards and Regulations:** All cite standards and one quarter regulations.
- Cluster 2 Competition and Standards:** All cite reasons related to competition and about a third standards.
- Cluster 3 Skills and Competition:** All cite skills, about a quarter cite competition and an insignificant number standards.
- Cluster 4 HR, Skills and Competition:** All cite human resources, and over one third cite both skills and competition, a minor number also cite standards.
- Cluster 5 Regulations:** All cite regulations, all other reasons are cited but in insignificant proportions.

The binary logit estimates are presented in table 8. The models are based on 1479 cases, with the proportion of correct predictions ranging from 0.69 to 0.91 and pseudo R^2 s ranging from 0.15 to 0.21. Only industry type is a significant driver for all clusters. All potential drivers listed in section 3.1 are important for at least one combined reason.

Table 8 Logit Parameter Estimates: Reasons for using Nationally Recognised Training

Variable	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Standards & Regulations		Competition & Standards		Skills & Competition		HR, Skills & Competition		Regulations	
	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio
Constant	0.090	0.15	-0.546	-0.65	-1.339*	-2.45	-1.111	-1.27	-3.864*	-4.84
Industry Type (Control: Professional & Other Services)	0.000*		0.002*		0.000*		0.000*		0.000*	
Agriculture, Forestry and Fishing	-3.279*	-5.87	0.394	0.76	0.809	1.75	0.758	1.01	0.550	1.53
Mining	-0.888	-0.90	-0.668	-0.46	-0.122	-0.12	2.827*	3.00	-1.324	-1.38
Manufacturing	-1.442*	-3.32	0.448	0.84	-0.748	-1.39	1.667*	2.34	0.104	0.28
Electricity, Gas and Water Supply	-3.291*	-2.61	0.790	1.32	-0.181	-0.26	-0.553	-0.45	0.602	1.32
Construction	-1.181*	-3.11	-0.353	-0.69	0.664	1.50	0.357	0.49	0.355	1.06
Wholesale Trade	-1.168	-1.66	-0.975	-0.82	2.791*	4.70	0.874	0.97	-3.098*	-2.74
Retail Trade	-0.385	-1.02	0.124	0.24	1.012*	2.26	1.015	1.44	-0.848*	-2.38
Accommodation, Cafes and Restaurants	-1.088*	-2.04	-0.824	-1.08	0.546	1.02	1.713*	2.23	0.034	0.08
Transport and Storage	-2.608*	-3.98	1.635	1.74	0.590	1.07	1.891*	2.45	0.701	1.62
Communication Services	-1.337	-1.47	0.248	0.28	-0.156	-0.17	0.524	0.48	0.868	1.45
Finance and Insurance	-0.994*	-2.37	-1.677*	-2.44	0.495	1.10	0.609	0.86	0.749*	2.11
Property and Business Services	-1.589*	-3.82	-0.718	-1.27	1.171*	2.65	2.148*	3.12	-0.890*	-2.52
Government Administration and Defence	-1.788	-1.82	-0.388	-0.36	0.631	0.74	1.806	1.84	0.398	0.41
Education	-1.216*	-2.02	0.846	1.29	0.087	0.15	1.079	1.32	0.575	1.06
Health and Community Services	-0.933*	-2.14	0.264	0.46	0.259	0.54	1.843*	2.62	-0.286	-0.73
Cultural and Recreational Services	-1.574*	-2.95	0.050	0.08	1.573*	3.35	-0.711	-0.78	0.255	1.58
Organisational Status (Control: Other Gov't Org'n)									0.036*	
Private 'for Profit'									0.371	0.59
Private 'Not for Profit'									-0.356	-0.56
Gov't Business Enterprise									-0.491	-0.42
Employee Size (Control: large)			0.056		0.000*				0.000*	
Small			-0.778*	-2.14	-0.311	-1.04			0.882*	3.14
Medium			-0.867*	-2.37	0.458	1.56			0.167*	0.59
Permanency of Employees (Control: High)	0.000*						0.010*			
Low	1.635*	5.52					-2.440*	-3.00		
Medium	0.697*	3.67					-0.098	-0.57		
Occupational Type										
Blue Collar Organisation	1.719*	6.66	0.562	1.87			-0.834*	-2.76	-0.629*	-3.38
White Collar Organisation	1.192*	3.88	0.732	1.89			0.972*	3.95	-1.632*	-6.55
Knowledge Worker Organisation	0.750*	3.55	0.635*	2.58			-0.707*	-3.57		
Strategic and Skill Variables										
Registered Training Organisation			-0.751	-1.70	1.042*	4.38			-1.350*	-4.23
Existence of a Business Plan			1.918*	3.95	-2.052*	-2.48				
Staff Training as part of Business Plan	1.589*	4.90	-2.119*	-4.78	1.732*	2.13				
Importance of Training	-0.312*	-2.58					0.388*	2.96		
Current Skill Level							0.326*	2.19		
Recruitment Difficulties			-0.382*	-3.26	0.206*	2.61	0.431*	4.54	-0.218*	-3.07
Nagelkerke R ²	0.175		0.152		0.149		0.228		0.212	
Proportion of Correct Predictions	0.860		0.911		0.791		0.833		0.688	

Significance levels are presented for retained multiple group categorical variables next to controls. * denotes significance at a 5% level.

Reasons for using Unaccredited Training

The results of the cluster analysis for unaccredited training are presented in table 9.

Table 9 Cluster Analysis: Reasons for using Unaccredited Training

Reasons	Cluster					Total
	1	2	3	4	5	
Skills	654 (100)	246 (44)	256 (41)	223 (49)	86 (25)	1465
Competition		561 (100)	267 (43)	144 (31)	48 (14)	1020
Standards			623 (100)	103 (22)	80 (24)	806
Responsive Workforce				459 (100)	17 (5)	476
Regulations					340 (100)	340
Number of Employers	654	561	623	459	340	

Numbers in parentheses represent the percent of employers in the cluster who cite the stated reason. Entries are **bolded** if cited by 50% or more of employers.

We provide the following interpretation of the identified clusters.

Cluster 1 Skills: All cite skills as the only reason.

Cluster 2 Competition and Skills: All cite competitive reasons and just under half also cite skill related reasons.

Cluster 3 Standards, Competition and Skills: All cite standards and over 40% cite both competition and skills as reasons.

Cluster 4 Responsive Workplace and Skills: All cite reasons related to a responsive workplace and nearly half cite skills. Competition and standards are also cited but by less than one third of employers.

Cluster 5 Regulations: All cite regulations, while all other reasons are cited but by a quarter or fewer of employers.

The binary logit estimates are presented in table 10. The models are based on 2637 cases, with the proportion of correct predictions ranging from 0.78 to 0.89 and pseudo R^2 s ranging from 0.07 to 0.12. Only industry type is a significant driver for all clusters. All potential drivers listed in section 3.1 are important for at least one combined reason.

Table 10 Logit Parameter Estimates: Reasons for using Unaccredited Training

Variable	Cluster 1 Skills		Cluster 2 Competition & Skills		Cluster 3 Standards, Competition & Skills		Cluster 4 Responsive Workplace & Skills		Cluster 5 Regulations	
	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio	B	T-ratio
Constant	0.323	0.64	-4.511*	-6.17	-1.780*	-2.97	-1.590*	-3.31	-4.359*	-7.46
Industry Type (Control: Professional & Other Services)	0.000*		0.000*		0.000*		0.000*			
Agriculture, Forestry and Fishing	0.325	1.14	-0.453	-1.27	-1.266*	-4.12	0.201	0.60	-0.135	-0.37
Mining	-0.550	-0.66	-0.569	-0.57	-0.615	-0.89	-0.845	-0.90	1.601*	2.55
Manufacturing	0.026	0.09	0.018	0.06	-0.438	-1.79	-0.580	-1.75	-0.265	-0.75
Electricity, Gas and Water Supply	-2.552*	-2.24	0.106	0.25	-2.934*	-3.13	-2.971	-1.82	1.118*	3.18
Construction	-0.056	-0.21	-0.584	-1.95	-0.690*	-2.85	-0.236	-0.76	0.435	1.35
Wholesale Trade	1.132*	3.71	0.541	1.58	-1.585*	-4.24	-0.153	-0.41	-0.874	-1.78
Retail Trade	0.430	1.65	0.276	0.98	-0.680*	-2.89	-0.302	-0.99	-0.576	-1.65
Accommodation, Cafes and Restaurants	0.050	0.15	0.567	1.62	-0.888*	-2.85	-0.213	-0.56	-1.145*	-2.11
Transport and Storage	0.153	0.45	-0.222	-0.50	0.025	0.08	-1.271*	-2.60	0.684	1.78
Communication Services	0.442	1.12	0.814*	2.09	-2.454*	-3.60	-0.374	-0.73	0.397	0.82
Finance and Insurance	-0.333	-1.10	0.056	0.18	-0.425	-1.70	0.077	0.24	0.728*	2.06
Property and Business Services	-0.314	-1.11	-0.231	-0.78	-0.171	-0.72	0.377	1.25	-1.608*	-3.40
Government Administration and Defence	-1.441	-1.71	0.771	0.87	-0.254	-0.30	0.566	0.80	-0.032	-0.03
Education	-0.783	-1.71	0.372	0.79	-0.080	-0.21	-0.896	-1.57	0.816	1.87
Health and Community Services	0.051	0.18	0.036	0.12	-0.896*	-3.45	0.407	1.31	0.411	1.19
Cultural and Recreational Services	-0.680	-1.88	0.530	1.59	-1.191*	-3.81	0.172	0.49	0.214	0.55
Organisational Status (Control: Other Gov't Organisation)	0.000*		0.068		0.095					
Private 'for Profit'	-1.553*	-4.09	0.681	1.27	1.005	1.95				
Private 'Not for Profit'	-1.591*	-3.99	0.358	0.66	1.223*	2.36				
Gov't Business Enterprise	-2.573*	-2.56	1.707*	2.23	1.309	1.72				
Employee Size (Control: large)			0.019*		0.000*					
Small			0.831*	2.30	-0.362	-1.45				
Medium			0.615	1.70	0.073	0.29				
Permanency of Employees (Control: High)			0.001*				0.099			
Low			0.661*	3.65			-0.439*	-2.13		
Medium			0.128	1.09			-0.022	-0.18		
Occupational Type										
Blue Collar Organisation			-1.291*	-5.47	-0.332*	-2.11	0.474*	2.80	0.692*	4.05
White Collar Organisation	0.368*	2.10					-0.372	-1.76	-1.366*	-4.11
Knowledge Worker Organisation			0.528*	4.63			-0.418*	-3.14		
Strategic and Skill Variables										
Registered Training Organisation	0.952*	5.57	-1.313*	-4.10	-0.411*	-1.98				
Existence of a Business Plan	0.457*	2.23								
Staff Training as part of Business Plan	-0.804*	-4.39	0.419*	3.10			0.827*	5.27	0.634*	3.73
Importance of Training			-0.228*	-3.26						
Current Skill Level	-0.218*	-2.73					0.290*	3.05	-0.279*	-2.56
Recruitment Difficulties							0.200*	3.13	-0.239*	-3.14
Nagelkerke R ²	0.082		0.106		0.065		0.080		0.117	
Proportion of Correct Predictions	0.781		0.813		0.793		0.847		0.887	

Significance levels are presented for retained multiple group categorical variables next to controls.

* denotes significance at a 5% level.

Modelling the Training choice between Nationally Recognised Training and Unaccredited Training

Modelling Approach

The SEUV data set provides information on the use of nationally recognised training (NRT) and unaccredited training. We will model the drivers for why NRT is used (either alone or in conjunction with other training types) versus when unaccredited training is used but without NRT.

The variables to be investigated as potentially important drivers for the choice of NRT against unaccredited training are categorised into three groups (organisational characteristics, strategic and skill recruitment) and are the same as described before (see section 3.1) for modelling the cited reasons for choosing training types.

The dependent variable is a binary categorical variable without any natural order. As consequence the binary logit model will be employed to develop an appropriate model and identify any important drivers for the choice of training type. In this case both the forward and backward stepwise procedures produced the same model.

Drivers of choice of NRT over Unaccredited Training

The binary logit model was based on 2965 cases with 1184 employers using NRT alone or with other training and 1781 employers using unaccredited training but without NRT. The model resulted in a pseudo $R^2 = 0.17$ with 67.9% of the two groups correctly predicted. Table 11 presents the logit estimates for the choice of NRT over unaccredited training. The majority of potential drivers were retained in the stepwise procedure, the omitted potentially important drivers were: organisational status, and blue and white collar organisations.

Table 11 Logit Parameter Estimates: NRT vs Unaccredited Training

Variable	B	T-Ratio
Constant	-2.065	-5.01
Industry Type (Control: Professional and Other Services)	0.000*	
Agriculture, Forestry and Fishing	0.533*	2.09
Mining	0.542	0.89
Manufacturing	-0.496*	-1.99
Electricity, Gas and Water Supply	1.737*	4.39
Construction	0.634*	2.72
Wholesale Trade	-1.015*	-3.21
Retail Trade	-0.457*	-1.96
Accommodation, Cafes and Restaurants	-0.399	-1.35
Transport and Storage	-0.233	-0.79
Communication Services	-0.950*	-2.49
Finance and Insurance	0.721*	2.94
Property and Business Services	0.363	1.53
Government Administration and Defence	0.588	0.83
Education	0.138	0.40
Health and Community Services	-0.474	-1.91
Cultural and Recreational Services	0.060	0.22
Employee Size (Control: large)	0.000*	
Small	-1.931*	-7.26
Medium	-1.543*	-5.80
Permanency of Employees (Control: High)	0.040*	
Low	-0.365*	-2.39
Medium	0.028	0.29
Occupational Type		
Knowledge Worker Organisation	-0.593*	-6.31
Strategic and Skill Variables		
Registered Training Organisation	0.457*	2.67
Existence of Business Plan	-0.873*	-3.53
Staff Training as part of Business Plan	1.105*	4.80
Importance of Training	0.130*	2.17
Current Skill Level	0.402*	5.51
Recruitment Difficulties	0.187*	3.74

Unaccredited Training without NRT is the reference group. * denotes significance at a 5% level.

Conclusion

This report has provided the technical details for modelling the drivers of the cited reasons for choosing particular training types and for the observed choice of NRT over unaccredited training. The report has described issues pertaining to data, and the analytical tools of cluster analysis and the binary logit model.

References

- Adena, M, 2006, *Response Bias and Weighting for the 2005 SEUV*, Australian Bureau of Statistics, Canberra.
- Chiu, T, Fang, D, Chen, J, Wang, Y, & Jeris, C 2001, 'A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment', *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2001*, 263-268.
- Hair, JE, Black, WC, Babin, BJ, Anderson, RE, & Tatham, RL 2006, *Multivariate Data Analysis*. 6th ed., Pearson International.
- Long, JS 1997, *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications: Thousand Oaks.
- National Centre for Vocational Education Research 2006, *Employers' Use and Views of the VET System 2005: Summary*,. NCVER, Adelaide.
- SPSS 2001, *The SPSS TwoStep Cluster Component*, White Paper - Technical Report, SPSS Inc.