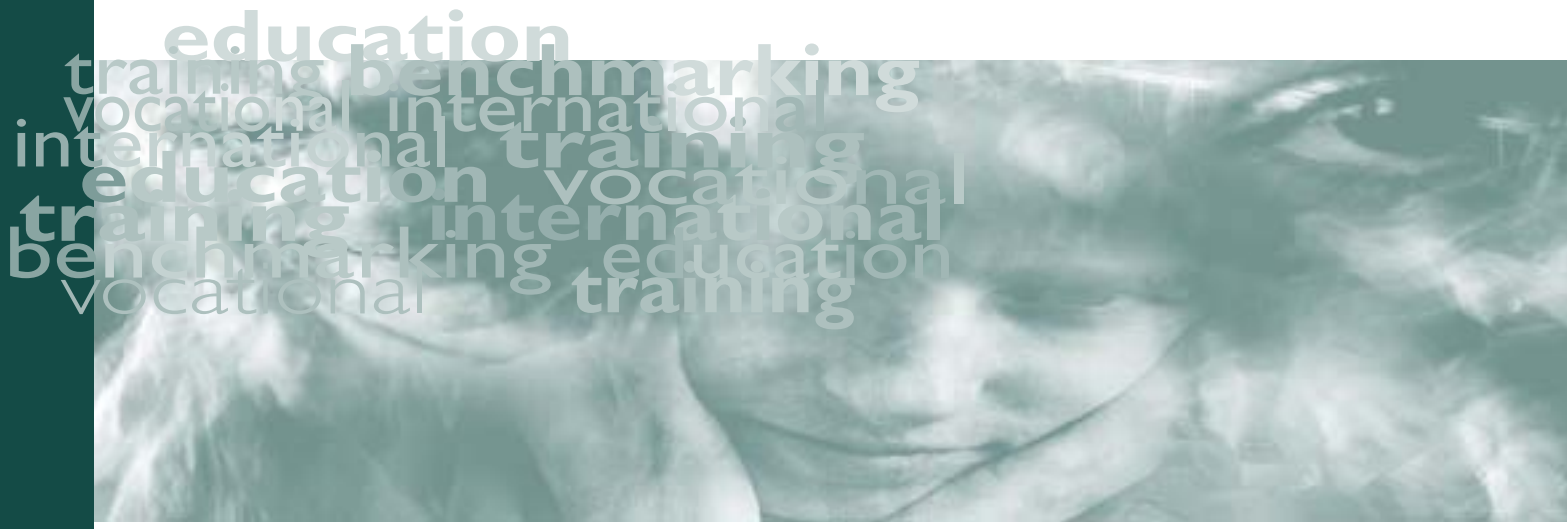
NCVER




# International benchmarking of vocational education and training

*Tim Wyatt*

NCVER

# International benchmarking of vocational education and training

*Tim Wyatt*

*Erebus Consulting Partners*

ISBN 1 920895 24 8 web edition

TD/TNC 76.16

Published by
NCVER Ltd, Level 11, 33 King William St, Adelaide, South Australia 5000
PO Box 8288, Station Arcade SA 5000
ph +61 8 8230 8400 fax +61 8 8212 3436
<http://www.ncver.edu.au>
ABN 87 007 967 311

# Contents

# Tables

# Acknowledgements

This project has undergone a number of changes in direction since its initiation. As the project progressed, the purpose of the project and the scope of what it could achieve have been significantly revised.

I would like to acknowledge the contribution of the late Dr Jeanne Griffith to the development of this report. It was originally envisaged that Dr Griffith would be a member of the research team for the project. As a former Associate Commissioner of the United States National Center for Education Statistics, and latterly a consultant to the Organisation for Economic Co-operation and Development (OECD), she has made a significant contribution to the field of international education statistics. Her advice in exploring the issues discussed in this paper was greatly appreciated. Unfortunately, as her health failed, she was not able to see the project through to the end.

I would also like to acknowledge Mr John Watt, of Clarinda Consulting, who contributed to research into benchmarking in the public sector, and Ms Leone Robb, who contributed significantly to our early attempts to develop case studies for selected countries. While this work has ultimately not been included in this report, it nonetheless has been important in shaping understanding of the issues which need to be resolved if we are to better understand the performance of Australia's vocational education and training (VET) system by comparison to those of other countries.

# Executive summary

## Aim

The aim of this project is to present a number of issues in relation to benchmarking the performance of the Australian vocational education and training (VET) system against that of other countries. It is not a project which attempts to *provide* benchmark data, but inform *about* the benchmarking process.

The report critically examines two approaches to benchmarking VET. The first of these approaches is that offered by the collection and publication of performance indicator data; the second approach is that represented by comparative case studies.

## What is benchmarking?

While comparative studies have long been a source of influence on educational policy-making in Australia, benchmarking as a specific form of study is a much more recent phenomenon. Underpinning the concept of benchmarking is an understanding of how organisations achieve a certain level of performance and the subsequent application of these approaches to another organisation.

The term 'benchmarking' is used to describe a large variety of different measurement and evaluation technologies which have been collected with one single aim: the improvement of organisational performance. Some define benchmarking as a technique similar to process mapping (for example, analysing work processes and comparing them to 'best practice'). Others see benchmarking as the activity of comparing outcomes or results of similar organisations.

Benchmarking generally involves the measurement of key performance criteria, identification of entities which may have similar performance data, comparison of the performance of the organisations, and analysis of the reasons for the differences in performance. Benchmarking as a tool for organisational improvement offers a range of benefits for organisations. These include:

- ✧ changing the culture of organisations from being inward-looking to being outward-looking
- ✧ improving the quantity and quality of performance information within an organisation
- ✧ making monitoring of agency performance by executive government and other stakeholders easier, thus improving accountability.

Benchmarking involves far more than the ad hoc collection of statistics. Benchmarking is a purposeful activity in which the processes used by an agency to deliver its products and services are compared with similar processes elsewhere. It is an ongoing and systematic process to search for and then introduce best practice into an organisation. It is by understanding the reasons for differences or gaps in performance that many organisations improve or grow.

The lessons from international experience are clear: benchmarking is a participative exercise, which needs to be well focused and sensitive to the needs of the organisations and stakeholders

involved. Benchmarking is therefore something which is not readily applied to an organisation (or activity), but something which is applied by organisations.

In examining the differences in the performance of the VET sector in different countries, it is natural to attempt to identify the factors which have promoted a higher level of performance. Are the differences in performance due to differences in legislative requirements, structural organisation, resource levels, demographic profile and so on, or are they due to less tangible factors such as motivation of students or morale of the workforce, competing demands from other forms of education and training, or the 'efficiency' of teaching?

The VET sector in Australia is, of course, not one organisation, but a complex set of public and private institutions and service providers. Some are institution-based, while others are based in the workplace. It would be wrong to speak of the Australian VET sector as a single entity, and there may well be as much variation within the VET sector in Australia as between Australia and other countries.

# Benchmarking through performance indicators

The first method for benchmarking the performance of Australia's vocational education and training system considered in this project draws on the experience of international organisations in developing performance indicators. The particular example used to illustrate this approach is the Organisation for Economic Co-operation and Development (OECD) publication, *Education at a glance* (OECD 1998, 2000a).

What distinguishes indicators from sets of statistics is not only the deliberate focus on issues of policy relevance, but also the emphasis placed on conceptual coherence. Considerable effort in the international indicator development process has been devoted to identifying the most appropriate model for organising how the indicators 'fit together'.

Assessing organisational or activity performance is usually not done using only one indicator; rather, a whole suite of indicators is necessary to give a more comprehensive picture of organisational performance. In assessing performance against this suite of indicators, the inter-relationships between the indicators need to be taken into account.

## Practical issues for benchmarking in the VET sector

The experience gained in developing the indicators in *Education at a glance* has highlighted a number of key practical issues which have emerged and which need to be considered in developing education benchmarks (particularly in a cross-national context), including problems of definition, data availability, and data quality. There are clearly significant gaps in international data sets at this time in regard to VET. The most serious of these is that the OECD data classifications do not have a separate category for VET.

Given the potential influence of these kinds of data, we need to identify the consequences which arise if the data are 'wrong'. What would the policy consequences be if we continually over-estimate our completion rates from VET, or under-estimate funding levels, relative to other countries?

# The comparative case study approach

A second approach to benchmarking makes use of what might be termed 'comparative case studies' to highlight similarities and differences between the performance of the education and training systems in different countries. The comparative case study approach gives priority to the analysis of circumstances (and changes in them) within countries rather than between them.

Instead of looking for correlations between standardised indicators in VET, qualitative data are used to identify trends within each country.

The case study approach allows a more thorough analysis of particular issues and can take greater account of the complexity and diversity of countries' institutional structures. This approach also accommodates the complex linkages between education, the labour market and other social institutions. The conceptual framework for the case study approach should therefore provide a tool which is useful for a more comprehensive understanding of an individual country's VET system and associated cross-national comparisons. The case study approach also has the practical advantage over the indicators approach in that it is less constrained by the varying coverage and availability of statistical data for each country.

More importantly, this approach provides valuable insights into how each system works. The individual country analyses enrich formal accounts of systems. They may challenge a country's image of its education system as open or flexible, or may draw attention to pathways which have more symbolic than actual substance.

There are three ways in which we can learn more by examining case studies alongside each other—that is, by making the study a comparative one. We can learn more about each individual country, we can learn from the similarities in the experience of different countries and we can learn from the differences between countries. Hypotheses developed on the basis of the comparisons can be put to a double test: can they explain differences between countries, and can they explain trends and patterns within individual countries?

The case study approach is by its nature, a time-consuming and therefore expensive exercise. Case studies cannot be developed effectively from a distance. They demand the participation of those with significant experience within the VET system of each country to locate, explain and contextualise appropriate data. Unless the case studies are constructed with a specific purpose in mind, it is possible that the data examined will have no relevance to the issue at hand. The usefulness of the case study approach can be enhanced by more clearly focusing on issues of particular policy relevance, rather than on simplistic descriptive overviews of national systems.

# Conclusions

An examination of the various approaches to benchmarking enables identification of the ways in which each type of data might usefully contribute to a better understanding of Australia's VET system. While it is clear that there is a range of significant issues attached to each approach in terms of the availability of data, further attempts to initiate serious benchmarking studies would appear to be worthwhile.

The analysis would suggest that contributing to the international comparative indicator efforts of organisations such as the OECD should continue to be supported. However, while they are necessary for the implementation of effective benchmarking, this level of data is not sufficient to be useful in driving improvement.

There is still much work to be done in developing the vocational and training statistical collections, both in Australia and other countries, in terms of developing common definitions, common language to describe the systems and common criteria for judging best practice and so on. However, we need to question whether making simplistic comparisons always makes good sense, particularly if we know that the data are flawed. We also need to avoid drawing conclusions about causes and effects where the data do not support this kind of analysis. The assumption that reforms successfully implemented in one country can be easily transplanted into another also needs to be challenged. Structural and cultural conditions can combine to provide a different context for reform in different countries.

Developing sets of indicators which have a deliberate policy focus may provide a more strategic approach to benchmarking and thus highlight Australia's international standing in relation to VET. To make benchmarking most useful, there is a need to ensure that individual benchmarking studies are clearly linked to the Australian VET policy agenda, and answer specific research needs.

# Introduction

## Aim

The aim of this project is to research and document issues related to benchmarking the performance of the Australian vocational education and training (VET) system against that of other countries. It is not a project which attempts to *provide* benchmark data, but inform *about* benchmarking. In this way our knowledge about the processes which can be used to more usefully inform policy and program improvement can be advanced. International comparisons can provide contextual information against which the performance of a particular country can be judged. They can help to provide answers to questions about those policies, structures and outcomes which determine 'best practice', and thus identify areas for improvement.

However, while there have been high expectations for quality improvement in the VET sector through international comparative studies, the evidence of any direct benefits in the Australian context is difficult to identify. We need to ask whether this is a problem which results from the design of individual studies, or is a problem inherent in the approaches taken. To answer this question involves examining both the theoretical advantages and disadvantages of various approaches, as well as examining some examples of their practical application.

This report critically examines two approaches to benchmarking VET. The first of these approaches is that offered by the collection and publication of performance indicator data (for example, Freeland 2000), in this instance best represented by the Organisation for Economic Co-operation and Development (OECD) international indicator series, as published in *Education at a glance* (OECD 1998, 2000a). The second approach to benchmarking is that represented by a form of comparative case study technique, as represented in the OECD (2000b) study.

The study draws on the available research and empirical evidence and identifies key initiatives which may be of interest and relevance to the VET sector in Australia. The focus of the research is on the possibilities and problems inherent in the methods of benchmarking, not the particular benchmarks provided in the examples given.

## The context for benchmarking in the VET sector

International comparisons have become part of the vocabulary of educational politics in Australia (Teese 1988). In their report on youth and work in Australia, the OECD examiners argued that 'enrolment in all post-compulsory education in Australia has always been low compared to enrolment in most other OECD countries' (OECD 1986, pp.27–8). OECD comparisons have been used to argue the need for increasing retention rates in schools and to expand funding for post-school study. Comparisons have been seen as a useful tool in formulating policy, at least within limits (Commonwealth Schools Commission 1987).

More recently, the Australian National Training Authority (ANTA) Ministerial Council specified the goal of the national VET system as '… the creation and maintenance of a national pool of skilled Australians sufficient to support internationally competitive commerce and industry' (ANTA 1998).

To support this goal, there is a need for a range of information about Australia's standing in terms of the pool of skills, as well as its international competitiveness.

At present, there is a limited amount of readily accessible information in the public domain which would allow simple and valid international comparisons of the performance of VET systems. Available data are often highly problematic. The annual national report on vocational education and training in Australia has relied on one simple international comparison, 'highest completed level of education (percentage of the labour force aged 25–64)', as a measure of stocks of vocational education and training skills against desired levels (ANTA 1998 p.20, 2000, p.52). The conclusion drawn from this indicator is that there may be a gap between the current Australian skill base compared with other countries. This conclusion, and the validity of the indicator as a measure of the stock of VET skills, has been hotly disputed by Australian systemic authorities.

Moreover, enumeration of the 'highest qualification obtained' is used as a proxy measure of potential work skills in the population. However, this measure distinguishes poorly between general and higher education and VET (despite the recent inclusion of a non-university tertiary category), and between recently obtained and outdated qualifications, or between qualifications used in the workplace and those which may be redundant. There is a need to identify data which more realistically and validly reflect the intention of the indicator. The relationship between qualifications and work skills, for example, requires further clarification. These problems are raised here as examples of the complexities which emerge when drawing conclusions from international benchmarking data. (This is not presented as an analysis of this issue, which is taken up later in this paper.)

However, this example illustrates the extent to which international comparisons have entered into the policy-making arena. More particularly, it illustrates the need for the development of sound conceptual and technical bases for indicators and benchmarks if measures such as these are to drive changes in education and training policy-making in positive directions. Debate about educational data is healthy and the possibility that it might be misused or misinterpreted by others should never be used as an excuse for not collecting or reporting it. Unless conducted and resourced properly, simplistic benchmarking can potentially do more harm than good.

# Benchmarking

## What is benchmarking?

While comparative studies have long been a source of influence on educational policy-making in Australia (for example, Shears & Mathews 1983), benchmarking as a specific form of study is a much more recent phenomena. Benchmarking originated in the private sector and focused on process improvement, typically in industrial situations, with the aim of giving an individual company a competitive advantage in the marketplace. Understanding how other organisations are able to achieve a certain level of performance and then incorporating these ideas into one's own organisation is the underpinning philosophy of benchmarking (Department of Finance and Administration 1996).

The term 'benchmarking' is used to describe a large variety of different measurement and evaluation technologies which have been bought together with one single aim: the improvement of organisational performance. Some define benchmarking as a technique similar to process mapping (for example, analysing work processes and comparing them to 'best practice'). Others see benchmarking as an activity whereby outcomes or results of similar organisations are compared. It is interesting to note that the first ANTA annual national report on VET (ANTA 1995) was titled *Benchmarking VET: The performance of the vocational education and training in 1995*. This report was ground-breaking, because it was the first time that such a collection of information had been brought together in a single report.

In considering what benchmarking might mean in a public sector setting, the Australian Government's Management Improvement Advisory Committee saw benchmarking as an 'on-going, systematic process to search for and introduce best practice into an organisation' (Management Improvement Advisory Committee 1994).

While benchmarking can allow comparison of the performance of different parts of a large organisation, the Department of Finance and Administration (1996) noted that: 'While benchmarking variations in performance between different parts of the same organisation may lead to incremental changes in efficiency or effectiveness, some authors and recent APS [Australian Public Service] experience suggests that the greatest performance gains from benchmarking are likely to be realised from external comparisons'.

The literature describes some common approaches to benchmarking the performance of the public sector. These approaches involve comparisons of processes, results and standards.

- *Process benchmarking* aims to improve the outputs of an organisation. It has a focus on the various work procedures in current use and aims to improve their quality and/or quantity. Because public sector agencies have typically been focused on process, such an approach has a great deal of appeal and an immediate broad practicality. Further, good processes clearly contribute to good outcomes for an organisation.
- *Results benchmarking* involves reviewing 'organisational outcomes against set outcome-related performance indicators' (Department of Finance and Administration 1996). This is typically achieved through the selection of a suite of performance indicators which best summarise the requisite outcomes for the organisation.

- *Standards benchmarking* involves 'setting a standard of performance which an effective organisation could be expected to achieve' (Cowper & Samuels 1997).

Trosa (1997) as Chair of the OECD Conference on Benchmarking, Evaluation and Strategic Management in the Public Sector indicates her belief that 'process benchmarking and results benchmarking are increasingly seen as complementary methods to be used in tandem'. This report adopts a similarly pragmatic view, in that to limit any potential line of inquiry would be counter-productive.

Regardless of the type of benchmarking attempted, there are some key questions to be answered in designing the process:

- How do we define and measure aspects of a program, activity or concept?
- What are the 'right' benchmarks?
- What are the most appropriate points of comparison?
- How will we know whether we have drawn the 'right' conclusions?

Accordingly, benchmarking generally involves the measurement of key performance criteria, identification of entities which may have similar performance data, comparison of the performance of the organisations and analysis of the reasons for the differences in performance.

Benchmarking and evaluation share many similarities in that they can both contribute to continuous improvement. Benchmarking, as simply a process of gathering and reporting data without an evaluative framework to 'make sense' of the data, is unlikely to be of much use to an organisation. At best, the data are simply a collection of more or less interesting 'facts', but with no context for understanding what the implications of those facts might be.

There are clearly a number of significant political and technical issues in undertaking benchmarking exercises. Some of these are discussed below. To be useful for driving improvement efforts, there will need to be a range of qualitative and quantitative data to enable an appreciation of the linkages between processes and results.

# Benefits of benchmarking

Benchmarking has been enthusiastically embraced by public and private sector organisations across the world, and its use is well documented in countries, including the United Kingdom (Cowper & Samuels 1997), Sweden (Dahlberg & Isakson 1997), and the United States. In Australia, the Commonwealth Department of Finance and Administration has been a leading advocate (see for example Department of Finance and Administration 1996), but in each of the state and territory jurisdictions there has also been significant activity in one form or another. Other organisations, such as the Productivity Commission in its *Report on government services* (Productivity Commission 2000), for example, has attempted to produce comprehensive benchmarking reports in several areas of government activity, both as a regular statistical series and through one-off special reports. Internationally, the OECD has actively promoted benchmarking as a means for understanding the social and economic status of member countries through the development of discussion papers, conferences and so on.

Benchmarking as a tool for organisational improvement offers a range of benefits for organisations. These include:

- changing the culture of organisations from being inward-looking to being outward-looking
- improving the quantity and quality of performance information within an organisation
- making monitoring of agency performance by executive government and other stakeholders easier, thus improving accountability

- ✧ the possibility that longitudinal comparisons in an individual organisation can be put in place, making tracking of improvements fairly straightforward
- ✧ cycles of continuous improvement becoming an integrated part of agency operations.

Furthermore, if both process and results benchmarking are pursued, there is a balanced approach to all aspects of organisational improvement (Department of Finance and Administration 1996).

The literature suggests benchmarking process 'can be effective at all levels of operation, from the conduct of individual processes, such as invoice handling, to the operational performance of organisations with tens of thousands of staff, such as a welfare benefits delivery agency' (Cowper & Samuels 1997).

This suggests that when we talk about 'improving the performance of an organisation', we need to be clear about what we mean by both 'improvement' and 'performance'. In terms of the Australian VET sector, we are not talking of improvement in some global qualitative sense on a single index, but improvement in particular attributes. In practice, because VET is complex, this will mean that multiple indicators will be needed to capture this complexity. Even in business, the notion that corporate performance can be measured simply by 'the bottom line' has been replaced by the notion of the need to understand the internal and external drivers of the business, as offered in the 'balanced scorecard' approach (Kaplan & Norton 1992).

# Success factors for benchmarking

Benchmarking involves far more than the ad hoc collection of statistics. In the current management context, benchmarking is a purposeful activity in which the processes used by an agency to deliver its products and services are compared with similar processes elsewhere. It is an ongoing and systematic process to search for and then introduce best practice into an organisation. It is by understanding the reasons for differences or gaps in performance that many organisations improve or grow.

> Benchmarking is not simply developing a scorecard of how your organisation measures up against others. It has to involve understanding the make-up of whatever business you are in, and how each element compares with others. (Bartos 1994)

In essence, the process is simple: it involves understanding the organisational processes of the organisation:

- ✧ identifying which of those processes are key to the successful delivery of the organisation's objectives
- ✧ establishing relevant performance indicators for those processes
- ✧ finding organisations with similar processes who perform better in terms of those indicators
- ✧ applying the lessons from those other organisations to your own processes.

In practice, the implementation of benchmarking is not as straightforward. Less than one-third of benchmarking studies in a range of industries actually lead to any performance improvement. Other recent quality improvement concepts such as total quality management (TQM) and business process re-engineering have had a similar lack of success (Haines 1996). It is fair to assume that not all benchmarking exercises are carried out in the same way. The Department of Finance and Administration (1996) has identified ten key factors related to the success of both process and results benchmarking. They are:

- ✧ involving all stakeholders
- ✧ understanding the organisation
- ✧ selecting processes important to clients

- developing appropriate performance indicators
- making appropriate comparisons
- recognising the link between results and process benchmarking
- finding a balance between balance and change
- introducing new incentives
- creating a new culture
- linking benchmarking with evaluation.

These factors are a timely reminder that benchmarking is not just a technical or statistical exercise. The lessons from international experience are clear: benchmarking is a participative exercise, which needs to be well focused and sensitive to the needs of the organisations and stakeholders involved. Further, there will be a number of iterations involved, until the right indicators have been developed, the right comparisons made and the right conclusions drawn. Benchmarking is therefore something that is not easily done to an organisation (or activity), but something which is done by organisations.

The VET sector in Australia is, of course, not one organisation, but a complex set of public and private institutions and service providers, some which are institution-based, others being based in the workplace. It would be wrong to speak of the Australian VET sector as a single entity, and there may well be as much variation within the VET sector in Australia as between Australia and other countries. How these principles, which have been developed in the context of comparing organisations or business activities, can be applied to an understanding of an educational sector is an important question taken up later in this report.

# Importance of understanding context

In examining the differences in the performance of the VET sector in different countries, it is natural to want to identify the factors which have promoted a higher level of performance. Are the differences in performance due to differences in legislative requirements, structural organisation, resource levels, demographic profile and so on, or are they due to less tangible factors, such as motivation of students or morale of the workforce, competing demands from other forms of education and training, or the 'efficiency' of teaching?

In discussing the issue of developing performance standards and undertaking benchmarking activities, there needs to be a clear understanding of local contextual policy and other issues to make sense of local situations and decisions. Accordingly, there is a need to understand:

- structural arrangements for education and training in each country, including review of new policy initiatives and developments
- governance and policy-making structures for VET in each country
- trends over time (for example, VET compared with general education)
- equity issues (participation by gender, racial minorities and socio-economic status where available)
- articulation between qualifications
- relationships between public and private providers
- information on employment outcomes
- information about the national priority given to VET (for example, expenditure on VET as a proportion of gross domestic product).

Developing this contextual understanding is itself far from easy. There is, for example, no single source of data which can be utilised to address each of the points above. The country education profiles developed by the National Office of Overseas Skills Recognition, Department of Education, Training and Youth Affairs (1992, 1993, 1995a, 1995b, 1997) are an attempt to compile comprehensive information cross-nationally, but they are considered to be contentious, incomplete, and too general to allow the kinds of analysis required. In addition, information sources tend to become out of date fairly quickly. Even within some countries, data on each of the above issues are rarely available from a single source, and if available at all, often require translation into English. Often such data have never been collated at a national level, and have to be compiled from regional sources where these are available.

The European Centre for the Development of Vocational Training (CEDEFOP) does compile and publish a wide range of information on recent developments in vocational education and training in European Union member states and some non-member countries. Of most relevance here are the country-specific reports, which provide descriptive information about aspects of the VET system (although in non-standard format). Some, but not all country reports provide statistical summaries of these aspects. These reports do not readily lend themselves to the production of composite tables which allow quick comparisons between countries. They are, however, useful for understanding the structural context of each country.

Despite the existence of these resources, to fully understand and describe how each country performs on each of the factors above would require a study in its own right. This emphasises the need for benchmarking to be conducted as a collaborative activity rather than as a desk exercise.

# Applying international benchmarking to the VET sector

## Introduction

Interest in international comparisons of education systems in recent years has been stimulated by globalisation of industry and the need to ensure that education systems are effectively contributing to the national economy.

In the United States for example, business leaders, policy-makers and researchers have expressed great interest in understanding how the United States education system compares with those of other countries. Australia has also participated actively in the development of stronger international educational data collection activities (see for example, Ruby 1988; Lokan 1999).

Decision-makers want to know whether students in Australia undertake similar levels of training and education, and are prepared for careers to the same extent as are students overseas. They also want to know whether curricula are as demanding and whether graduation requirements are as stringent abroad as they are at home. Also of interest is whether training systems incur similar costs (Freeland 2000).

Why is so much importance placed on international comparisons? Simply stated, understanding others helps us to better understand ourselves. International indicators provide an opportunity to compare our own performance with that of other countries, to identify similarities and differences between our system and others, and to suggest new approaches to the challenge of providing a world-class education.

Most of this comparative activity has, however, been focused on school-level education, although there has been some activity at the tertiary level. Very little activity has focused specifically on the VET sector. Whether any of these activities (such as the International Education Association Third International Mathematics and Science Study (Harmon et al. 1997), truly represents 'benchmarking' activities per se, as defined above, is questionable, but they do provide some data which can be used for benchmarking processes.

## Benchmarking process

The discussion above has indicated that, from a practical point of view, the benchmarking process involves a number of distinct stages.

### Identification of existing data sets

This first stage involves an analysis of the information collected by various organisations, in this case, education systems. Education systems typically generate large amounts of data, but these data may not necessarily be transformed into information which can be used for performance assessment of the system. Limitations in this regard arise from:

- ✧ statutory and regulatory reasons (for example, certain data use is limited by law, such as the performance of individual students in individual schools)

- technical reasons (for example, systems do not necessarily have the technical capacity in their information systems to make sense of all the data collected)
- political reasons (for example, there may be agreements between stakeholders that particular sorts of data will be only used in particular ways, often restricting the use and publication of these data)
- physical reasons (for example, data collected may be undertaken by a variety of different parts of an education system and these data are never aggregated into a useable form).

### Comparison of data sets and analyses of differences

This second stage involves an analysis of the differences between individual data sets. For substantive understandings to emerge from this analysis, it is necessary to have a good grasp of the context in which the original data were collected. For instance, there may be significant policy, structural, political, social, economic and historical differences which act as intervening variables for the purposes of analysis. Having understood these possible differences, and their effects on the data sets, then the analysis of differences can commence. Generally speaking, at this stage the aim is to try to identify where there are meaningful differences and to find explanations for these differences.

### Testing explanations for observed differences

Once differences are observed, it is necessary to explore why these differences arise. This generally means entering into a dialogue with the particular jurisdiction concerned to determine why this is so. It is this cooperative sharing of information, analysis and discussion of differences which is at the heart of successful benchmarking activities.

### Determining improvement

This stage involves determining the ability of a single system to take action to improve aspects of its operation. This can be a significant challenge in itself since the practices of other jurisdictions often need modification for introduction into another jurisdiction. Further, there can be significant limitations on changes to a system, including industrial agreements, finance, change fatigue and so forth.

### Implementing the required changes

This may involve a large number of staff and require such actions as an organisational change strategy, training and development, new curriculum materials, a changed organisational structure, changes to credentialling methods and so forth. The expectations for change resulting from benchmarking in vocational education and training need to be realistic. In particular, there is a need to acknowledge that it is both time- and resource-intensive. While it may lead to significant and enduring change, this is unlikely to happen immediately in the education sector.

## Benchmarking in the literature

While the rhetoric of the need for international comparative data is strong, a search of the VET literature reveals that there are very few examples of how benchmarking has been applied within the Australian VET sector, particularly as defined in the discussion above. Montague and Evans (1996) document a trial process benchmarking educational delivery in the New South Wales VET sector, designed to investigate whether teachers would find using process benchmarking useful for improving their practice. The Victorian Office of Technical and Further Education (1997) has produced a resource manual designed to assist organisations in the VET sector to conduct benchmark projects, to develop their benchmarking skills, and to provide a valuable tool to use in

their quality improvement activities. Other researchers, for example, Casey (1995) have discussed the role of benchmarks in ensuring best practice in VET, established by a combination of results from key performance indicators. None of these, however, is concerned with the publication of international benchmark data.

There are a few examples of specific-issue benchmarking available. One concerns benchmarking investments in employer-provided training and development across organisations and countries (American Society for Training and Development 2000). The data in this study, although useful for illustrative purposes, are drawn from a relatively small sample of organisations in each country relative to the total of all training providers. Stalker (1994) examines the issue of the impact of global migration on economies or labour forces most severely affected by the arrival, presence or departure of migrants. The information in this study is now very dated (using 1990 data), and is more relevant for understanding the context of VET provision than the performance of the sector. Cullen (1997) reports the results of a study into the benchmarking of the Australian qualifications profile in relation to work skills and international competitiveness; however, it should be noted that the methodology for the study and the conclusions drawn have been controversial. Sloan (1994) discusses the economic theory of training, and such issues as reasons for failure in the training market, with some international comparisons. The comparisons here are not central to the paper and are for illustration only.

The tradition of cross-national comparisons of VET appears to be stronger in Europe. Several examples are represented in the literature. Mulder (1997) compares school-to-work transition arrangements in Germany, the Netherlands, England and Wales, and the United States. Parkes (1997) makes a comparative analysis of education and training systems among various European Union countries, particularly France, Germany and the Netherlands in relation to post-16 training to explore the origins, meanings and effectiveness of strategies employed to separate learning, qualifications and learners into distinct types. Lasonen (1998) also analyses post-16 education reform in eight European upper secondary systems. Ryan (2000) also conducts a similar analysis in relation to apprenticeships in smaller European countries. These studies use the technique of comparative analysis, but do not rely heavily on statistical benchmarking to draw their conclusions.

The ANTA annual national reports, sometimes known as the ANTA benchmarking reports, present a range of data on key performance measures (KPMs) for the Australian VET system. These data include:

- *KPM 1:* Skill outputs produced annually within the domain of formally recognised VET:
  - participation in vocational education and training
  - skill outputs
  - load pass rates
  - successful completions by individual students
  - other successful outputs.
- *KPM 2:* Stocks of vocational education and training skills against desired levels
- *KPM 3:* Employer views on vocational education and training
- *KPM 4:* Student outcomes from vocational education and training
- *KPM 5:* VET participation, outputs and outcomes achieved by client groups
- *KPM 6:* Public expenditure per publicly funded output
- *KPM7:* Public expenditure per total recognised output
- *KPM 8*: Total expenditure on vocational education and training.

Apart from the indicator of 'highest completed level of education (percentage of the labour force aged 25–64)', discussed earlier, the 'benchmarking' of the data in these reports comes variously

from comparison of figures over time, between different Australian states and territories, between different sectors of the education system, and different kinds of providers. The report also recognises the importance of context in providing explanations of the data shown in the indicators. This kind of benchmarking, while largely not cross-national in nature, is similar to that used in international reports such as the OECD's *Education at a glance* discussed below.

There are undoubtedly many other individual reports and studies which could be cited and which take a cross-national view of aspects of VET systems. However, the fact that a study includes data about, or descriptions of, aspects of VET in different countries does not mean that they can necessarily qualify as an international 'benchmarking' study according to the definitions discussed earlier.

The following chapters examine two different approaches to benchmarking which appear to offer some prospect for further enlightenment.

# Benchmarking through performance indicators

## Background

The first method for benchmarking the performance of Australia's vocational education and training system considered in this project draws on the experience of international organisations in developing performance indicators. This approach has been the product of more than a decade of work to develop comparable measures relevant to educational policy-makers. The particular example used to illustrate this approach is the OECD publication, *Education at a glance*. While there are other international indicator efforts, Australian governments have enthusiastically supported the OECD effort since its inception (see for example, Ruby 1988). The introduction to *Education at a glance* describes the rationale for the project thus:

> Governments are seeking effective educational policies that enhance the social and economic prospects of individuals, contribute to economic productivity, provide incentives to promote the efficiency of the administration of schooling and help mobilise additional resources to meet increasing demands for education and learning. To inform the process of policy formation and to reinforce the public accountability of education systems, the OECD continuously seeks to develop indicators that can provide insight into the comparative functioning of education systems – focusing on the human and financial resources invested in education and on returns to those investments.
>
> A quantitative description of the functioning of education systems allows countries to see themselves in the light of other countries' performance. Through international comparisons, countries may come to recognise weaknesses in their own education systems, while also identifying strengths that can otherwise be ignored in the heat of domestic debate. The OECD education indicators show whether variations in educational experiences within a country are unique or if they mirror differences observed elsewhere. The OECD education indicators are the product of an ongoing process of conceptual development and data collection, the objective of which is to link a broad range of policy needs with the best internationally available data. (OECD 1998, p.5)

Performance indicators are conceptualised in the literature (for example, Oakes 1986) as concise, policy-relevant measures which allow monitoring of the health of the system. They do not attempt to measure everything which can potentially be measured, but attempt to describe the key features which categorise an enterprise.

What distinguishes indicators from sets of statistics is not only the deliberate focus on issues of policy relevance, but also the emphasis placed on conceptual coherence. Considerable effort in the international indicator development process has been devoted to identifying the most appropriate model for organising how the indicators 'fit together'. Most performance indicator systems address four major aspects of the functioning of organisations:

- *Input indicators:* which describe the resources available to particular education systems. This might involve staffing numbers, the value of monetary investments made in education, subsidies paid, capital works programs and so forth.
- *Workflow or process indicators:* which attempt to quantify the amount of work undertaken by a particular system. This might involve quantifying the numbers of students in a particular system, the number of courses run in particular systems and so forth.

- *Output indicators:* which explain system outputs, such as the number of students graduating from particular courses or systems.
- *Outcomes indicators:* which are based around the results generated from the system. In the case of vocational education and training, these can be considered as the ability to gain employment or further training, the ability to generate a 'good' income, or the ability to accumulate further qualifications which lead to a higher-level credential.

These four aspects are represented in the indicators presented in *Education at a glance*, which have been organised under six thematic headings:

- the demographic, social and economic context in which education systems operate
- the financial and human resources which countries invest in education, comparing the resources which countries invest in education, relative to national wealth; the number of students and the size of the public purse; the ways in which education systems are financed; and the sources from which the funds originate and the deployment of resources across different functional categories
- access to education, participation, progression and completion. Trends in enrolments in the various levels of education and types of educational institutions are shown to indicate how the supply and demand of educational resources have evolved in different countries
- the labour force participation of young people 15 to 29 years of age, both while in education and following the completion of initial education
- the learning environment and the various ways in which school systems are organised. It shows data on teacher compensation, demographics of the teaching force, the statutory time that teachers are required to teach and students required to be in a classroom, subject emphasis in the curriculum, how decision-making authority is distributed across levels of government and the use of computers in schools
- the individual, social and labour market outcomes of education.

Assessing organisational or activity performance is usually not done using only one indicator, rather, a whole suite of indicators is necessary to give a more comprehensive picture of organisational performance. In assessing performance against this suite of indicators, the inter-relationships between the indicators need to be taken into account. Early writers about performance indicators in education (for example, Oakes 1986; Porter 1988) place considerable emphasis on choosing indicators which reflect a model of how the education system is believed to 'work'.

# Practical issues for benchmarking in the VET sector

The experience gained in developing the indicators in *Education at a glance* is salutary and has been documented in a number of publications (for example, Wyatt 1994; Wyatt & Ruby 1988; Griffith 2000). From this process a number of key practical issues have emerged which need to be considered in developing education benchmarks (particularly in a cross-national context), including the following.

## Definition issues

Problems of definition take several forms, and pose the greatest threat to international comparisons.

Foremost are those which concern the construct validity of the proposed indicator—the extent to which the data accurately and adequately reflect the constructs of interest. For many of the constructs of interest in the VET sector only proxy measures are available. For example, in systems where completion is recorded on a 'met requirements'/'not met requirements' basis (that is, there is

no differentiation in the quality of graduates), there is a temptation to report on the quality of the system in terms of the quantity of graduates. The problem here is the assumption that the level of skills, knowledge and understanding of all graduates is the same or equivalent—a highly unlikely situation. Where completion is itself used as a measure of the quality or 'holding power' of the system, there may also be a faulty assumption that 'completion' is the goal of students enrolled in VET. Anecdotal evidence suggests that not all students intend to gain formal recognition for their studies, particularly those whose aspirations are non-vocational (such as for hobby courses, or those using VET study as a stepping stone to other forms of higher education). Some students may also wish to gain particular skills rather than a full qualification.

Second is the problem of aggregation of institutional and regional data to form national/system-level estimates, and the attendant decisions about inclusion or exclusion of certain data elements. For example, even within one system, there are often disagreements about what constitutes an 'enrolment'. Is it obtained by a head count of students on a particular census day, or when an application to enrol is received or an enrolment fee paid, on registration at the first class, or after attendance at a certain number of classes?

Similarly, there are differences in interpretation about what counts as a vocational education and training institution (consider the problems caused by multi-campus colleges, joint school–VET programs, work-based programs and industry training), what counts as a 'course' (number of classes, hours per week or year, articulation to other qualifications) and so on. How to aggregate costs is also an issue, particularly when these costs may be shared between various government agencies and/or employers, and may include direct spending and less direct subsidies or tax rebates.

None of the problems is without solution, but typically require considerable negotiation to develop protocols for measurement and establishment of common practices. The International Standard Classification of Education Definitions (ISCED) has been the standard means for attempting to obtain consistency between countries in OECD and United Nations Educational Scientific and Cultural Organisation (UNESCO) studies for many years. In spite of the fact that there is an international system for classifying levels of education, schools in different countries which are classified at the same ISCED level may not provide the same programs or have the same functions.

## Data availability

Even where there is common agreement about definitions, there is no certainty that any particular system or nation will have internal data collection systems to support any particular indicator for policy, cost or even historical reasons. Almost all attempts to develop international indicator series contain gaps for some countries on some indicators (or else are heavily qualified). Sometimes these gaps occur because data gathered at institutional level for local management purposes are not aggregated to systemic levels. At other times there is no agency responsible for collection/action on the particular indicator. For example, it is conceivable that comparative data on attendance levels in VET courses could be an indicator some systems may wish to benchmark, but this topic is covered in few national VET statistical reports. Breaks in statistical series are also common, for example, when courses are reclassified from one administrative grouping to another, or from one level to another.

## Data quality

When data are aggregated from multiple sources, there will always be issues relating to the quality of the data. In many instances the quality control mechanisms which are employed by agencies/systems are unknown. Most national educational data are collected either as by-products of administrative records collected at institutional levels, as census collections required by national authorities, typically for funding or accountability purposes, or by sample surveys.

Errors can occur at a number of stages and for a number of reasons. Sample surveys, often used to measure 'satisfaction' levels, for example, are often affected by non-response biases, and small changes in question wording can produce considerable change in the outcome of the survey. Administrative records are susceptible to idiosyncratic interpretation by local officials and teachers, duplications, omissions and transpositions in data recording, assignment of incorrect codes, and so on. Centrally mandated collections are also open to recording errors, but are also susceptible to fraud and misrepresentation, particularly when there is some financial incentive to inflate figures. 'Phantom' students are often found in audits of enrolment counts when, for example, funding is allocated on a per-head basis. Most central agencies responsible for statistical collections have established extensive audit procedures to detect, correct or at least estimate the extent of these problems, but it cannot be assumed that any data set is absolutely error free.

Comparisons of education finance are particularly attractive to policy-oriented audiences. Two international organisations, the OECD and UNESCO have long been active in compiling information on education expenditure, but the resulting statistics and indicators have been of limited value and often misleading because the underlying figures have not been internationally comparable. Countries were interpreting the international agencies' requests for expenditure data differently, making conflicting decisions about what to include in, or exclude from, their finance statistics, and categorising and measuring expenditures in idiosyncratic ways.

Comparability is possible when countries base their statistics on uniform concepts, equivalent categories and consistent operational definitions. While comparability is a matter of degree, comparability problems are of three main types, the first being *problems of scope or coverage,* which occur when countries differ with respect to which expenditure items are included in or excluded from statistics on education spending. These differences reflect conflicting definitions of the boundaries of education sectors, uneven coverage of institutions or funding sources, and inconsistent coverage of spending for particular aspects, services or items of expenditure. For example, significant discrepancies occur between countries which include nurse education as part of the non-university tertiary spending figures.

The second relates to *problems of categorisation* which occur when countries inconsistently classify expenditure items by level of education, type of service provider, nature or resource category, or source of funds. The contribution of private funds is a particular source of discrepancy here.

The last, *problems of measurement,* results from the use of incompatible methods to quantify the amount spent within a given expenditure category.

This is not an exhaustive discussion of the problems which can be encountered in cross-national benchmarking, but a sample of the challenges involved in undertaking this task. It should be noted that the OECD is now making significant efforts to develop more comparable indicators. It has been working with UNESCO and EUROSTAT since 1995 to jointly collect data.

> The return on this collaboration has been substantial improvement in the collection, organisation and quality of international education statistics, as well as a reduction in the time taken to publish the indicators. The continuing implementation of common definitions, the use of high standards for quality control and better data documentation have improved the international comparability of education statistics. (Griffith 2000)

It should be noted here that the Australian Vocational Education and Training Management Information Statistical Standard (AVETMISS) project involving all VET systems in Australia has attempted to provide greater consistency and reliability to the VET data collected and reported at state/territory and national level. Over a number of years, audit procedures in relation to student enrolment reporting, for example, have seen substantial improvement in the quality of these data sets and subsequent enhancement in the level of confidence in comparisons based on these data.

# Current relevant VET statistics

Current indicators published by the OECD which concern VET include the following:

- ✧ educational attainment of the adult population
- ✧ inter-generational change in completion of tertiary education
- ✧ expected number of years in employment, unemployment and time outside the labour market
- ✧ educational expenditure relative to gross domestic product
- ✧ government support for education as a share of total public expenditure
- ✧ educational expenditure per student
- ✧ overall participation in formal education
- ✧ participation in and completion of secondary education
- ✧ access to and participation in tertiary education
- ✧ completion of and drop-out from tertiary education
- ✧ participation in continuing education and training by adults
- ✧ education and work of the youth population
- ✧ expected years in education, employment and non-employment between the ages of 15 and 29
- ✧ the reasons for youth unemployment
- ✧ youth unemployment and employment by level of educational attainment
- ✧ labour force activity by level of educational attainment
- ✧ labour force activity of persons leaving education
- ✧ earnings and educational attainment
- ✧ private, fiscal and social rates of return to education.

It might be assumed at first glance that there is a rich source of comparative information from which a picture of the relative performance of the Australian VET sector can be assembled. Closer examination reveals that interpretation of many of these indicators is problematic.

Three indicators are reproduced below to illustrate the range of issues which limit the confidence which can be accorded to interpretation. Table 1 shows international data on 'highest education attainment of the labour force aged 25–64' (also reproduced in the ANTA annual national report). The figures suggest that in Australia, higher education dominates the non-university tertiary education sector in terms of post-school qualifications attained by the workforce. However, as noted in the ANTA annual report:

> Because of differences in data definitions and variations in collection methods between countries, it is difficult to measure the gap (positive or negative) between Australia's stock of skills, and that of other countries. Furthermore, international comparisons do not show differences between countries on unstructured skill acquisition at workplace, or through formal short training courses that do not lead to a qualification, or indeed on the broader picture of lifelong learning. (ANTA 2000)

Freeland (2000) discusses problems with the use of the 'country mean', as a basis for comparison. Because the structural, cultural and demographic differences between countries are great, the concept of an 'average' would appear to make little sense in this context.

**Table 1: International comparisons: Highest completed level of education—percentage of the labour force aged 25–64 years, 1998**

| Country | No post-compulsory | Less than upper secondary | Upper secondary education | Non-university tertiary | University-level education | Total post-compulsory education (incl. upper secondary) |
|---|---|---|---|---|---|---|
| Australia | 38 | 33 | 10 | 19 | 62 | **100** |
| Canada | 16 | 28 | 36 | 21 | 85 | **100** |
| New Zealand | 23 | 41 | 22 | 14 | 77 | **100** |
| Germany | 12 | 56 | 15 | 16 | 87 | **100** |
| Switzerland | 16 | 59 | 10 | 15 | 84 | **100** |
| United States | 11 | 51 | 9 | 29 | 89 | **100** |
| Country mean* | 32 | 42 | 10 | 16 | 68 | **100** |

Note: * The country mean includes the following countries which are not reported in the original table: Korea, Austria, Belgium, Greece, Hungary, Ireland, Iceland, Japan, Mexico, Luxembourg, Spain, Poland and Turkey.

Source: OECD (2000a, p.34)

Table 2 shows labour force participation rates (1999) by level of educational attainment and gender for populations 25 to 64 years of age. While the pattern of results is similar in Australia to the benchmark countries, and is also of roughly similar levels, the indicator is weak because the age range is great. It disguises potentially significant differences between younger and older workers.

**Table 2: Labour force participation rates (1999) by level of educational attainment and gender for populations 25 to 64 years of age**

| | | Below upper secondary education | Upper secondary and post-secondary non-tertiary education | Tertiary-type B | Tertiary-type A and advanced research programs | All levels of education |
|---|---|---|---|---|---|---|
| Australia | Men | 79 | 89 | 91 | 93 | **86** |
| | Women | 54 | 66 | 81 | 73 | **63** |
| Canada | Men | 74 | 88 | 91 | 90 | **86** |
| | Women | 48 | 73 | 80 | 84 | **72** |
| Germany | Men | 76 | 84 | 88 | 92 | **84** |
| | Women | 47 | 70 | 82 | 83 | **66** |
| Switzerland | Men | 91 | 94 | 96 | 97 | **94** |
| | Women | 63 | 74 | 88 | 81 | **73** |
| United States | Men | 74 | 87 | 90 | 92 | **87** |
| | Women | 50 | 72 | 82 | 81 | **73** |

Source: OECD (2000a)

Table 3 shows relative earnings of the population with income from employment by level of educational attainment. This table does not include the non-university tertiary category, but instead includes the 'tertiary-type B' category, which is roughly equivalent to the United States junior college level or associate degree level. It is not clear how it maps against the Australian Qualifications Framework (AQF) levels.

**Table 3: Relative earnings of the population with income from employment by level of educational attainment and gender for the populations 25 to 64 and 30 to 44 years of age (upper secondary education =100)**

| | | Below upper secondary education | | Tertiary-type B education | | Tertiary-type A and advanced research programs | | Tertiary education | |
|---|---|---|---|---|---|---|---|---|---|
| | | 25–64 | 30–44 | 25–64 | 30–44 | 25–64 | 30–44 | 25–64 | 30–44 |
| Australia | Men | 87 | 83 | 120 | 116 | 144 | 138 | 136 | 131 |
| | Women | 85 | 84 | 113 | 112 | 154 | 154 | 137 | 138 |
| | M+W | 79 | 75 | 103 | 101 | 136 | 131 | 124 | 120 |
| Canada | Men | 84 | 81 | 109 | 112 | 148 | 143 | 130 | 128 |
| | Women | 76 | 69 | 116 | 118 | 164 | 165 | 137 | 138 |
| | M+W | 83 | 79 | 106 | 109 | 152 | 149 | 128 | 128 |
| Germany | Men | 77 | 63 | 105 | 101 | 149 | 131 | 126 | 116 |
| | Women | 85 | 68 | 104 | 106 | 160 | 167 | 128 | 134 |
| | M+W | 78 | 62 | 106 | 104 | 157 | 144 | 130 | 123 |
| New Zealand | Men | 76 | 74 | x | x | x | x | 137 | 135 |
| | Women | 74 | 73 | x | x | x | x | 129 | 130 |
| | M+W | 76 | 74 | x | x | x | x | 136 | 136 |
| Switzerland | Men | 81 | 77 | 122 | 124 | 144 | 140 | 135 | 133 |
| | Women | 73 | 80 | 131 | 133 | 154 | 160 | 145 | 151 |
| | M+W | 75 | 76 | 140 | 142 | 161 | 157 | 153 | 151 |
| United States | Men | 65 | 63 | 119 | 123 | 183 | 180 | 176 | 173 |
| | Women | 63 | 65 | 120 | 120 | 170 | 177 | 163 | 170 |
| | M+W | 67 | 66 | 118 | 120 | 180 | 178 | 173 | 171 |

Note: x Data not applicable or included in another column of the table.

Source: OECD (2000a)

# Gaps in existing information

There are clearly significant gaps in international data sets at this time in regard to vocational education and training. The most serious of these is that the OECD data classifications do not have a separate category for VET. In some cases, VET is included as 'upper secondary', and in other cases, 'non-university tertiary', depending on the level of course studied and the country involved. Australian VET statistics are contained in both classifications, meaning that international comparisons based on OECD data are blurred at the margins (Freeland 2000).

# Limitations of existing data collections

Almost all of the writers who have contributed to the conceptual development of indicators have mentioned various problems and limitations. These problems can be classified broadly as those relating to the statistical or conceptual properties of the individual indicators, and those which relate to the interpretation and use of indicators. These problems can be summarised as follows:

- ✧ Indicators provide limited information.
- ✧ There can be problems with simple models.
- ✧ There are problems with the collection and analysis of indicator data.
- ✧ Indicators can deflect attention from real solutions to problems.
- ✧ Indicators create political pressures for sometimes inappropriate responses.

Various commentators examine these issues in greater depth. David (1988), Porter (1988) and Garbutcheon-Singh (1988) discuss some of the political implications of indicator systems, in terms of possible effects on local control over teaching, curriculum, and discouragement of innovation. They point to the problem of time and money being spent on collecting data which will then have limited use, and of indicators becoming ends in themselves.

Shavelson et al. (1989) and Murnane (1987) discuss the problems of operationalising indicator development and shaping conceptual frameworks. They point to the problem of operationalising measures of performance which are closely related to the concept of interest. There are always trade-offs to be made in determining the level of disaggregation for reporting measures. It is also difficult, especially in cross-national comparisions, to determine sub-group breakdowns likely to be helpful in explaining performance trends. Without these disaggregations it is extremely difficult, if not impossible, to fully understand the composition of particular results. For example, it is quite possible that two very different sets of circumstances could have created similar macro-indicator scores.

The problems faced in developing and interpreting education indicators are not dissimilar to those encountered in the development of economic indicators (see, for example, Horn & Winter 1990). Murnane and Pauly (1988) draw three lessons from this experience. First, it is important to develop multiple measures. No single indicator offers a workable basis for assessing the performance of the education and training system, just as no single measure—such as the unemployment rate—provides a reliable measure of the health of the economy. Second, we should keep the original purposes of indicators firmly in mind. As the term suggests, indicators indicate, they do not prove anything. They are intended to help us focus on more questions. They do not provide all of the answers, nor can they be expected to. Third, it is important to educate the users of indicators. In the economic field, a rich set of indicators has been available for nearly 50 years, and users have become well versed in piecing together a picture of economic performance by examining trends in more than one indicator. The development of a comparable set of education indicators is still evolving, and most users are still learning how to use them and to interpret the patterns appearing in the data (see Odden 1990, for example).

Further problems come when performance indicators are 'cobbled' together from information collected for other purposes. It is common, for example, for administrative data to be used for performance reporting, because of their apparent ready availability and low cost. What tends to be reported is what is easy to collect, rather than what *should* be collected. There is also rarely a good match between the technical requirements of the performance indicator and those of administrative collections. It is relatively easy, for example, to record teacher qualifications used administratively to determine salary levels. However, teacher qualifications are a poor proxy for *teacher quality,* which is the variable of interest in measuring the performance of the teaching process.

Avoiding overstatement of both long- and short-term benefits of indicators is also important. Although they can play a significant role in both the planning process and accountability, indicators are by no means the sole answer to an organisation's problems. They can at best only point to the existence of problems or successes, and cannot suggest solutions (Wyatt 1994).

The technical limitations to the construction and interpretation of indictors also need to be recognised. The literature on technical problems can be grouped in two main sub-categories: those addressing the issue of how to make fair comparisons (for example, Freeland 2000), and those concerned with the adequacy of available measures. How change can be measured over time, whether and how various measures should be adjusted to account for various influences (for example, whether expenditure should be adjusted for purchase price parity), and how to deal with the multi-level nature of educational processes, are all issues which require resolution in any international comparison. The sophisticated statistical techniques available for analysing indicator data lose much of their power if the basic data they require are inadequate.

# What evidence is there of the success of performance indicator comparisons?

Whether or not the results of international comparisons have been useful and relevant to their intended purpose is a legitimate question. Arguably, the proper role of international benchmarking is to:

- identify trends in how education and training policy and performance is changing internationally
- throw a critical light on taken-for-granted assumptions about how existing institutional and social arrangements operate
- suggest alternative approaches to attaining desired outcomes from VET (Freeland 2000).

If this is the case, then we might agree with Keep (1991) that 'international comparisons generate an abundance of results relevant to VET policy and have provided a pivotal role in defining the perceptions of problems to which a policy response is required'.

We know that research information enters into the policy-making process in subtle and indirect ways. If we were to ask, did publication of any particular indicator lead directly to an identifiable policy change, most often the answer would be 'no'. If we were to ask, did it contribute to the global picture or general weight of opinion, the answer might be different.

Data are sometimes used directly in program administration. Although this is probably their most visible use, it is probably one of the least frequent. Data are often used in policy planning and evaluation. In such cases, researchers' data analyses provide statistical information which helps policy-makers expand their knowledge base for decision-making. Such data analyses are intentionally targeted to influence specific policy debates. Data are also used in more indirect ways to influence policy through the conduct of basic policy research—work which is produced either inside or outside government in the interests of generally expanding knowledge and understanding of issues. Basic research influences public policy decision-making by identifying issues, presenting new insights to ongoing concerns, clarifying aspects of continuing debate, and drawing attention to old issues. The findings from such research frequently filter into public policy debates through media coverage and broad public discussion about emerging issues or insights provided into ongoing issues. The international comparisons commonly published in OECD reports are regularly used by lobby groups to support calls for changes to government policies.

Given the potential influence of these kinds of data, we need to identify the potential consequences if the data are 'wrong'. What would the policy consequences be if we continually over-estimate our completion rates from VET, or under-estimate funding levels, relative to other countries?

As the discussion above suggests, these real and potential inaccuracies in OECD indicator data come not only from poor record-keeping at institutional and systemic levels or inadequate sampling designs, but also from structural differences in how certain data elements are defined. How private expenditure on training is treated is an issue which has not as yet been satisfactorily resolved. The inclusion of nurse education, and in some cases, teacher education in vocational education statistics, is another case in point. In some countries, nurse education is undertaken at tertiary institutions (universities) resulting in degree qualifications, but in other countries, formal nurse training more clearly aligns with the post-school, non-tertiary category with a higher on-the-job training component. Inclusion or exclusion of these students has a significant impact on apparent participation levels.

These structural differences can be accommodated in subsequent re-analyses, provided the basis on which the original data and the assumptions underpinning them are known (but unfortunately

often they are not). To be fair to the OECD, it must be acknowledged that this organisation is continually refining its methods for developing indicators which will provide a genuine indication of performance relative to policy objectives. It has, from the beginning of its publication of international indicators, provided caveats and explanations in recognition of the difficulty of making valid international comparisons. Unfortunately, these caveats are often overlooked or ignored by authors of reports or users of international comparative statistics. The OECD is also continually refining its guidelines for valid and feasible data collection.

In fact, Griffith (2000) argued that the development of both national and international databases is one of the most important outcomes of the project to develop international education indicators. The entire process of networks, advisory groups, special study groups, survey development and publication production has been one which has been highly effective as a learning exercise. A number of countries have actually used the International Education Indicators Study (INES) system to stimulate development of a national education data system. Some countries, typically those with highly decentralised education systems, had little or no national data on the process or performance of their systems. Over a period of time, a number of these countries have built national data systems able to participate in this process and to compare their country's performance in a meaningful way. The United States, Canada, Spain, and Switzerland have all published reports modelled on *Education at a glance* which compare provinces or states within their own country.

The key point here is not whether any particular indicator comparison is accurate or not, but whether this approach can in fact ever lead to improvement on the issue in question. Many benchmarking exercises appear to be premised on the assumption that if the Australian system were to be more like the benchmark systems, then our performance or competitiveness would improve. This assumption ignores the importance of context in producing the outcome, and ignores the finding that high performance on one benchmark may preclude high performance on another. Whenever a small set of countries is selected for comparison, there will always be the problem that inappropriate countries against which to benchmark, are chosen. The weakness of this approach is similar to the 'excellence trap' (see Cuttance 1994). Much of the work on change and development in organisations undertaken in the 1980s led to prescriptive statements about the characteristics of high-performance organisations, for example, *In search of excellence* (Peters & Waterman 1982). The 'trap' inherent in this approach was that it led others to believe that by attempting to implement and develop or mimic these features in their own organisations, they would also achieve high performance. However, what is appropriate for one education system at a particular stage of its development, is not necessarily appropriate for others.

# How can benchmarking based on performance indicators be improved?

Improving the quality of benchmarking using performance indicators requires the adoption of a more strategic approach to the selection of the indicators to answer more specific policy-related questions. It involves a far more purposive collection of data for analytical rather than descriptive purposes. Sets of indicators based on some theoretical model of the inputs, processes and outcomes believed to characterise a VET system are unlikely to be useful in this context. This strategic approach aligns more closely with the definition and use of the term 'benchmarking' as used in industry described earlier in this report. It goes well beyond the simple assemblage of statistics or even sets of statistics on a particular theme and beyond 'indicators', where the principal purpose is to raise further questions, in that it seeks answers to specific policy questions. The performance indicators developed in this conceptualisation thus seek confirmation or rejection of a set of hypotheses on the basis of evidence gathered from a set of benchmarking 'partners' (in this case, a group of countries), to reach a set of policy conclusions.

A key element of this approach is the development of a framework within which the evaluative comparison is made. This framework summarises the key questions used to evaluate particular aspects of the VET system. Benchmarking can be performed at firstly, an aggregate level, for positioning one country and exploring the validity of relationships, and secondly, at the benchmark level, which focuses on selected countries and seeks to examine the processes which contribute to the overall position, to test specific hypotheses and explore issues. This approach thus attempts to combine the quantitative data from international statistical data collections with qualitative data from case studies to improve the explanatory power of both. The comparative statistical data assist in providing a wider context for the data from any particular country, with the qualitative data offering insights into the reasons why particular circumstances may have arisen.

# The comparative case study approach

## Introduction

A second approach to benchmarking uses what might be termed 'comparative case studies' to highlight similarities and differences between the performance of the education and training systems in different countries.

The comparative case study approach gives priority to the analysis of circumstances (and changes in them) within countries rather than between them. Rather than looking for correlations between standardised indicators in VET, qualitative data are used to identify trends within each country. Comparisons across countries are made at a more qualitative level and related to a conceptual framework.

The case study approach allows a more thorough analysis of particular issues and can take greater account of the complexity and diversity of countries' institutional structures. Allowance can also be made for the complex linkages between education, the labour market and other social institutions. The conceptual framework for the case study approach should therefore provide a tool which is useful for more effectively understanding a single country's VET system as well as for cross-national comparisons. The case study approach also has the practical advantage over the indicators approach in that it is less constrained by the varying coverage and availability of statistical data for each country.

More importantly, this approach provides valuable insights into how each system works. The individual country analyses augment the formal accounts of systems. They may challenge a country's image of its education system as open or flexible, or may draw attention to pathways which have more symbolic than actual substance.

The comparative case study approach is illustrated in Ball et al.'s (2001) study which explores aspects of adult retraining and re-skilling in Australia and Korea. The study considers the context for adult education in the two countries—the structural arrangements for the delivery of adult re-skilling and retraining. The authors also consider briefly the outcomes of such training. This analysis is summarised in table 4. From this analysis, the benefits and limitations of the approaches used in the two countries are identified and the policy implications for each country discussed.

The study concludes that effective comparisons between the Australian and Korean general and vocational education and training systems must consider the impact of differing social, economic and cultural factors.

> Notwithstanding these differences, the vocational education and training systems in both countries need to be responsive to economic changes of a similar nature occurring in both economies. These changes include the growth in the service sector of the economy that is taking place in both countries at the expense of the manufacturing sector; changes in the nature of work, including growth in non-standard employment such as part-time and casual work and outsourcing; and increasingly, the need for generic skills in the workplace, such as total quality management and teamwork. Increasingly, both countries will need to give more attention to adult retraining and re-skilling as the proportion of young people in the working age population declines. Employers' skill requirements, currently met through training young people, will by necessity need to be met by re-skilling older workers. In addition, as the

working population ages, a higher proportion of the workforce will need training to upgrade their skills.

Given the similar forces of change operating in the two economies, Australia, in responding and adapting to these changes, can learn from recent Korean experience. Traditionally, adult education and training in Korea has been predominantly the domain of enterprises. The growth of the large industrial conglomerates meant that employees would remain with the one employer for a long period of time. This work culture meant that the larger employers were virtually assured of recouping the benefits of any investment made in the training of employees. (Ball et al. 2000, p.89)

The conclusion from this comparison is that nurturing linkages between training providers and industry in both countries is essential to the provision of adequate and relevant skill development for students.

**Table 4: Comparison of adult retraining and re-skilling in Australia and Korea**

| | Australia | Korea |
|---|---|---|
| *VET system* | | |
| Initial training | Apprenticeship and traineeship system or vocational education and training course leading to an AQF qualification. | Initial training is provided at high schools and two-year junior colleges under the formal education system, and at vocational training centers as non-formal education. |
| Relationships between initial training and further training (retraining and reskilling) | Linkage between initial training and further training (retraining and reskilling) is facilitated by the AQF. Adults can upgrade their skills by studying for a higher level of AQF qualification. Alternatively, they can enrol in a module or group of modules to acquire specific skills. | Initial training and further training are not closely linked. Further training is provided at firms with financial support from employers. Initial training emphasises generic knowledge and skills to learn new skills. Adults are encouraged to continue studying in post-secondary educational institutes. |
| *Adult retraining and reskilling* | | |
| Accredited training course | System allows multiple-entry points. Adults can access initial training through pre-vocational courses and enabling courses (basic language, literacy and numeracy courses) prior to enrolling in a course leading to an AQF qualification. | The *Act on promoting workers training* encourages adult workers to take training with financial support from the Employment Insurance Fund. Courses available include generic skills as well as specific skills. |
| Non-accredited training course—privately funded | Short specific-purpose courses provided by employers for a particular student group. | Short specific-purpose courses are provided at private training institutes. |
| Training contents | Training packages developed by industry. Competency-based training focusing on specific skills. | Courses are available in core-competencies that teach generic skills. Training programs developed by industry. Job upgrading training focusing on specific skills. Courses are provided to train generic skills including corporate culture. |

Source: Adapted from Ball et al. (2001)

# An illustrative example

The study by the OECD, *From initial education to working life* (OECD 2000b) provides an illustration of the application of this approach. The pathways approach uses the metaphor of travel to analyse education and training systems. It presents systems as networks of interconnected pathways, which may vary in how they are structured and in the nature of their interconnections (OECD 2000b). The case study data from this study are not reproduced here, for reasons of space, but table 5 provides an example of the kinds of data which could be developed for this approach.

Such data, while limited, at one level can be used to reflect on the outputs of various educational pathways, at least in gross terms. Their weakness is that they do not capture the transfers or flows between pathways. In any case, it is doubtful whether data on flows as well as stocks (that is, participation rates) can be reported in a single table. As with the performance indicator approach, the case study approach requires consideration of multiple sets of data.

**Table 5: Persons who have completed one or more educational qualifications—highest level of qualification (Australia)**

| All persons | 15–19 | 20–24 | 25–34 | 35–44 | 45–54 | 55–64 | 15–64 |
|---|---|---|---|---|---|---|---|
| *Post-school qualification* | *88.8* | *97.8* | *97.7* | *96.9* | *97.6* | *97.9* | *97.3* |
| Higher degree | 0.0 | 0.5 | 2.4 | 4.7 | 5.8 | 5.0 | 3.7 |
| Post-graduate diploma | 0.0 | 1.0 | 3.5 | 5.8 | 6.0 | 4.8 | 4.5 |
| Bachelor degree | 0.5 | 25.9 | 24.7 | 21.0 | 18.1 | 12.3 | 21.0 |
| Undergraduate diploma | 5.9 | 6.0 | 8.5 | 11.5 | 12.7 | 14.2 | 10.4 |
| Associate diploma | 2.7 | 11.9 | 8.7 | 6.3 | 6.9 | 5.3 | 7.6 |
| Skilled vocational qualifications | 11.7 | 19.5 | 23.1 | 22.7 | 23.1 | 30.6 | 23.0 |
| Basic vocational qualifications | 67.5 | 30.0 | 23.8 | 21.8 | 21.6 | 20.6 | 23.8 |
| Level not stated or inadequately described | 0.5 | 3.0 | 3.0 | 3.2 | 3.3 | 5.2 | 3.3 |
| Other qualification | 1.2 | 2.2 | 2.3 | 3.1 | 2.4 | 2.1 | 2.6 |
| Certificate of less than one semester | 5.0 | 1.3 | 1.5 | 2.1 | 1.3 | 1.1 | 1.6 |
| Secondary school certificate | 6.1 | 0.9 | 0.9 | 1.0 | 1.2 | 1.0 | 1.0 |
| **Total** | **100.0** | **100.0** | **100.0** | **100.0** | **100.0** | **100.0** | **100.0** |

Source: Unpublished data from the ABS Education and Training Experience Survey, 1997

Policy-makers have attempted to influence participation in VET by changing the nature and structure of the pathways available. They use a number of policy levers to try to increase the attractiveness of VET study, including:

- ✧ increasing the linkages between school and VET, and VET and university courses
- ✧ increasing the flexibility of transition between VET options
- ✧ elevating the status of VET qualifications
- ✧ punishing non-participants by withdrawal of welfare benefits
- ✧ mandating VET credentials as pre-requisites for vocational employment or registration, and so on.

Policy changes based on the concept of pathways can be observed in a wide range of countries (OECD 2000b). It would thus seem that there would be something to gain from a closer examination of the experience of these attempts to influence participation in VET. However, the very familiarity of the metaphor of pathways may make it a blunt instrument of analysis and research. We all know what we mean by pathways—or think we do—until we apply the concept to a cross-national study, and find that we all use the term in different ways. A major task for the study was to develop a conceptual framework which would be clear and rigorous enough to apply to countries with very different education and training systems.

In this conceptual framework, education is provided in the form of programs, which may be linked together in sequences to form pathways. Each program may also form part of other pathways with different starting points and destinations. An education system is a network of pathways. In some systems, the different pathways may be separate, with few opportunities to move between them; in

others there may be numerous crossroads, junctions and intersections, so that any program may be entered by students travelling along a diversity of pathways.

We can therefore describe an education system and its components in terms of the characteristics of:

- *Individual programs:* their entrance requirements, content, level, learning site, duration, certification, and the access or entitlements to further education or training which they might give.
- *Individual pathways:* their starting point, length, sequence, and destination; and in terms of the characteristics of their component programs (for example, general or workplace).
- *Systems of pathways:* the age or stage at which pathways diverge, the number and diversity of pathways, their interconnectedness and so on. A system may be described in terms of the characteristics of the average or typical pathways within it.

There are two recurring difficulties in applying the framework within countries. The first arises from the limitations of available statistics. The second is the problem of how empirically to identify pathways, independently from policy intentions and from the journeys which students make along them.

# Limitations and benefits of the approach

One of the benefits of the comparative approach is that it alerts us to the complexity and even idiosyncrasy of each country's pathways. There may be as much variation within systems as there is between them. For example, a country may have flexible pathways connecting certain programs, but it may have restricted pathways in other parts of the system. Seeing the different systems described alongside each other helps us to resist the temptation to classify any one country using an over-simplified typology.

The comparative case study approach allows an empirical benchmarking in that each country can compare its own experience with that of others. This too must be done cautiously, as the range of countries included is typically small. The danger is that 'best practice' may not be present in the group of countries selected for comparative analysis.

The performance of the VET sector can only be properly considered in the context of the entire educational effort of the country, including secondary, tertiary and community education. Indeed, the blurring of the boundaries between sectors represents one of the most substantial challenges to interpreting and understanding systemic performance. Some argue that a country's education and training system must be understood in relation to other social institutions such as its labour market and economy, its industrial relations system and its system of government (Maurice, Sellier & Sylvestre 1986). The societal perspective is consistent with the case study approach. However, it warns us against explanations which focus on a single set of factors such as pathways or the labour market, and it warns us against using simple causal models to study the complexities of education systems and their place in society.

While the case study method is largely qualitative in nature, this does not mean that the country descriptions are data-free. In fact, in order to develop profiles on particular themes, it is imperative that the profile be built from credible within-country data systems. The difference between the resulting profiles and those in the indicators model is that the statistical base does not have to be as constrained by the demands of cross-national comparability. Understanding VET student pathways, for example, makes demanding requirements of statistical data. Ideally, it requires data which:

- describe flows as well as stocks, or levels of participation
- describe flows through the whole education and training system, including apprenticeship and part-time education, and into the labour market
- compare students from different educational and social backgrounds
- describe the system at different points in time, so that comparison of trends is possible.

In the 1998 OECD study, the data for most countries do not satisfy all of these criteria. Countries typically have better data for stocks (that is, levels of participation) than they do for flows (that is, transitions between programs), and even where they do exist, series tend to be incomplete. In some countries, flow data are available only from one-off studies, or for particular regions, rather than on a regular national basis. In many cases, the available flow data describe particular transition points rather than flows throughout the whole system. Transitions between different full-time programs are well described in most countries, but part-time study, apprenticeships, entry into the labour market, and breaks in study are less adequately covered. Flows across more than one transition point are also not well described. Most countries routinely record the gender of their students in their national statistics. However, information on other student characteristics, such as social class, ethnicity or nationality, tends only to be available from ad hoc research studies.

While countries vary widely with respect to the adequacy and coverage of the available data, within each country, the data need careful interpretation in the light of the specific national context. It would therefore be very difficult to make direct statistical comparisons of flows and pathways across countries or to construct indicators to represent the key concepts of the pathways perspective. For the time being, at least, OECD 2000b concludes that it is more appropriate to proceed on a case study basis, and to analyse each country within the constraints of its available data. By piecing together a wide variety of types of evidence we can start to construct an explanation for changes in each country, and the pathways perspective provides a conceptual framework for doing so.

There are three ways in which we can learn more by putting case studies alongside each other—that is, by making the study a comparative one. We can learn more about each individual country, we can learn from the similarities in the experience of different countries and we can learn from the differences among countries. Hypotheses formed on the basis of the comparisons can be put to a double test: can they explain differences between countries, and can they explain trends and patterns within individual countries?

The case study approach is, by its nature, a time-consuming and therefore expensive exercise. It demands the participation of those with significant experience within the VET system of each country, to locate, explain and contextualise appropriate data. The case study approach is thus one which cannot be undertaken lightly.

While the OECD study used as an example here did focus on one important issue (pathways through the system), the question of how best to define what might be compared requires some further exploration. It is not possible, nor should it be attempted, to compare every single facet of VET systems from a single case study (no matter how comprehensive). At the other end of the spectrum, unless the case studies are constructed with a specific purpose in mind, it is possible that the data examined will have no relevance for the issue at hand.

# Conclusions

This section draws together conclusions based on the information collected as part of this study and reflects on options for future directions which might usefully be pursued in the immediate future.

The purpose for examining the different approaches to benchmarking considered here has not been to further the qualitative versus quantitative debate, but to point to ways in which types of data might usefully contribute to a better understanding of Australia's vocational education and training system. While it is clear that there are several significant issues attached to each approach in terms of the availability of data for both Australian and overseas VET systems, further attempts to initiate serious benchmarking studies would appear to be worthwhile.

The analysis above would suggest that contributing to the international comparative indicator efforts of organisations such as the OECD should continue to be supported. However, while they are necessary for effective benchmarking to take place, this level of data is not in and of itself sufficient to be useful in driving improvement.

There is still much work to be done in developing the vocational and training statistical collections, both in Australia and other countries, in terms of developing common definitions, common language to describe the systems and common criteria for judging best practice and so on. From a statistical point of view, we know far more about stocks than we do flows through the system; we know more about the quantity of activity than the quality of activity or outcomes; and we know very little about the interaction of internal and external processes.

We need to question whether making simplistic comparisons always makes good sense, particularly if we know that the data are flawed. We also need to avoid drawing conclusions about causes and effects where the data do not support this kind of analysis. The assumption that reforms successfully implemented in one country can be easily transplanted into another also needs to be challenged. Structural and cultural conditions can combine to provide a different context for reform in different countries.

Developing sets of indicators which have a deliberate policy focus may provide a more strategic approach to benchmarking and thus illuminate Australia's international standing in relation to VET. To make benchmarking most useful, there is a need to ensure that individual benchmarking studies are clearly linked to the Australian VET policy agenda, and answer specific research needs. Benchmarking studies should not be 'fishing expeditions', where the data are trawled to see what of interest might be revealed. Benchmarking studies should address real issues and the needs of the system, leading to better policy formation.

# References


American Society for Training and Development 2000, *ASTD international comparisons report: ASTD's annual accounting of worldwide patterns in employer-provided training*, ASTD, Alexandria, VA.

ANTA (Australian National Training Authority) 1995, *Benchmarking VET: The performance of the vocational education and training in 1995,* ANTA, Brisbane.

—— 1998, *Annual national report on vocational education and training, 1997, vol. 3,* ANTA, Brisbane.

—— 2000, *Annual national report on vocational education and training, 1999, vol. 3,* ANTA, Brisbane.

Ball, K, Young-Hyun, Lee, Oanh Phan & Young-Sun Ra 2001, *Adult retraining and reskilling in Australia and South Korea,* NCVER, Adelaide.

Bartos, S 1994, 'Benchmarking—serving the public', paper given at the 1994 Australasian Evaluation Society Conference, 21–23 September, Canberra Convention Centre, Canberra.

Casey, D 1995, 'Best practice in benchmarking: A module to support the development of best practice in the vocational education and training sector', loose leaf folder, Best practice and quality assurance research package, Regency Institute of TAFE, Port Adelaide.

Commonwealth Schools Commission 1987, *In the national interest: Secondary education and youth policy in Australia,* Commonwealth Schools Commission, Canberra.

Cowper, J & Samuels, M 1997, 'Performance benchmarking in the public sector: The United Kingdom experience', paper presented at a conference on international benchmarking, Danish Ministry of Finance, 20–21 February, Copenhagen.

Cullen, R B 1997, *Workskills and national competitiveness: A benchmarking framework,* ANTA, Brisbane.

Cuttance, P 1994, 'Quality systems for the performance development cycle of schools', paper prepared for the International Conference for School Effectiveness and Improvement, Melbourne, January.

Dahlberg, L & Isaksson, C 1997, 'The implementation of benchmarking from a Swedish perspective', paper presented at a conference on international benchmarking, Danish Ministry of Finance, 20–21 February, Copenhagen.

David, J L 1988, 'The use of indicators by school districts: Aid or threat to improvement?', *Phi Delta Kappan,* vol.69, no.7, pp. 449–50.

Department of Finance and Administration 1996, *Measuring up: A primer for benchmarking in the Australian public service,* discussion paper no.4, Resource Management Improvement Branch, Canberra.

Freeland, B 2000, *International comparisons of vocational education and training,* NCVER, Adelaide.

Garbutcheon-Singh, M 1988, 'Indicators or vindicators of quality', *Equity Links,* vol.36, p.18.

Griffith, J 2000, 'INES—A story of progress', remarks prepared for the Fifth General Assembly Tokyo, Japan, 11–13 September, unpublished paper.

Haines, G 1996, 'Don't tinker—transform', *Management,* September, p.11.

Harmon, M, Smith, T A, Martin, M, Kelly, D L, Beaton, A, Mullis, I, Gonzalez, E & Orpwood, G 1997, *Performance assessment in IEA's Third International Mathematics and Science Study,* <http://isc.bc.edu/timss1995i/PAreport.html>, viewed May 2004.

Horn, R & Winter, C 1990, 'Common factors in the development of economic indicators: Lessons learned', paper for the National Center for Education Statistics (NCES) Indicators Panel, Washington, DC, October.

Kaplan, R & Norton, D 1992, 'The balanced scorecard: Measures that drive performance', *Harvard Business Review,* January.

Keep, E 1991, 'The grass looks greener: Some thoughts on the influence of comparative vocational training research on the UK policy debate', in *International comparison of vocational education and training for intermediate skills,* ed. P Ryan, Falmer Press, London.

Lasonen, J 1998, 'Improving the quality of vocational education through international collaboration and comparisons', paper presented at International Vocational Education and Training Association Conference 1998, *Quality matters in international vocational education and training,* Ankara, Turkey, September.

Lokan, J 1989, 'PISA: Definition and development of assessment frameworks', paper presented at the conference, *International comparisons of pupil performance—issues and policy,* Leeds, United Kingdom.

Management Improvement Advisory Committee 1994, *Benchmarking: Improving performance in the APS,* AGPS, Canberra.

Montague, A & Evans, P 1996, *Benchmarking by teachers: Report of a study trialling process benchmarking in vocational education and training,* NSW TAFE Commission, Sydney.
Maurice, M, Sellier, F & Sylvestre, J J 1986, *The social foundations of industrial power,* MIT Press, London.
Mulder, M 1997, 'School-to-work transition: Cross-national comparison of approaches', *International Journal of Vocational Educationa and training,* vol.5, no.1, pp.69–84.
Murnane, R J 1987, 'Improving education indicators and economic indicators: The same problems', *Educational Evaluation and Policy Analysis,* vol.9, no.2, pp.101–16.
Murnane, R J & Pauley, E W 1988, 'Lessons from comparing educational and economic indicators', *Phi Delta Kappan,* vol.69, no.7, pp.509–13.
National Office of Overseas Skills Recognition, Department of Education, Training and Youth Affairs 1992, *Country education profiles – Switzerland,* AGPS, Canberra.
—— 1993, *Country education profiles –Canada,* AGPS, Canberra.
—— 1995a, *Country education profiles – Australia,* AGPS, Canberra.
—— 1995b, *Country education profiles – New Zealand,* AGPS, Canberra.
—— 1997, *Country education profiles – Israel,* AGPS, Canberra.
Oakes J 1986, *Education indicators: A guide for policy makers,* RAND Publications, Santa Monica, CA.
Odden, A 1990, 'Making sense of education indicators: The missing ingredient', in *Education indicators for quality, accountability and better practice,* eds T Wyatt & A Ruby, Australian Conference of Directors-General of Education, Sydney.
OECD (Organisation for Economic Co-operation and Development) 1986, *Youth and work in Australia: Comprehensive policy agenda,* Manpower and Social Affairs Committee, Review of Youth Policies in Australia, OECD, Paris.
—— 1994, *Making education count: Developing and using international indicators,* OECD, Paris.
—— 1998, *Education at a glance,* OECD, Paris.
——2000a, *Education at a glance,* OECD, Paris.
——2000b, *Fom initial education to working life: Making transitions work,* OECD, Paris.
Office of Training and Further Education 1997, 'How to benchmark in the VET sector', loose-leaf folder, OTFE, Melbourne.
Parkes, D 1997, 'The international perspective: Learning from international comparisons', in *Qualifications for the future: A study of tripartite and other divisions in post-16 education and training,* eds G Stanton and W Richardson, Further Education Development Agency report, vol.2, no.5, FEDA, London.
Peters, T & Waterman, R 1982, *In search of excellence,* Harper and Rowe, New York.
Porter, A 1988, 'Indicators: Objective data or political tool?', *Phi Delta Kappan,* vol.69, no.7, pp.503–8.
Productivity Commission 2000, *Report on government services,* vol.1, AGPS, Canberra.
Ruby, A 1988, 'The international indicators project', in *Indicators in education,* eds T Wyatt & A Ruby, Conference of Directors General of Education, Sydney.
Ryan, P 2000, 'Institutional requirements of apprenticeships: Evidence from smaller EU countries', *International Journal of Training and Development,* vol.4, no.1, pp.42–65.
Shavelson, R, McDonnell, L, Oakes, J & Carey, N 1989, *Indicator systems for monitoring mathematics and science education: A sourcebook,* Rand Corporation, Santa Monica, CA.
Shears, L & Mathews, J 1983, *Youth policies: A report to the Minister of Education,* Victoria, AGPS, Melbourne.
Sloan, J 1994, 'Research into the costs and benefits of VET', paper presented to Research Priorities in Vocational Education and Training: A Discussion, 20–22 April, ANTA, Brisbane.
Stalker, P 1994, *Work of strangers: A survey of international labour migration,* International Labor Organisation, Geneva.
Teese, R 1988, 'Australian education in cross-national perspective: A comparative analysis with France', *Comparative Education,* vol 24, no.3, pp.305–16.
Trosa, S 1997, *Remarks from the chair, OECD Conference on Benchmarking, Evaluation and Strategic Management in the Public Sector,* OECD, Paris.
United States National Centre for Education Statistics 1998, *The condition of education indicator report,* NCES, Washington.
Wyatt, T J 1994, 'Education indicators: A review of the literature', in *Making education count: Developing and using international indicators,* OECD, Paris.
Wyatt, T & Ruby, A 1988, *Indicators in education,* Conference of Directors General of Education, Sydney.


The National Centre for Vocational Education Research is Australia's primary research and development organisation in the field of vocational education and training.

NCVER undertakes and manages research programs and monitors the performance of Australia's training system.

NCVER provides a range of information aimed at improving the quality of training at all levels.

ISBN 1 920895 24 8 web edition