

To Capture Production or Well-being? A Review Article on *Towards Measuring the Volume Output of Education and Health Services: A Handbook*

Aled ab Iorwerth¹
Finance Canada

ABSTRACT

This review article evaluates the report *Towards Measuring the Volume Output of Education and Health Services: A Handbook* produced by the OECD. Traditionally, input costs have been used to estimate the value of education and health services. The *Handbook* provides detailed analysis and recommendations regarding appropriate methodologies and data to improve volume output measures for these services. The author welcomes the efforts of the OECD in this area. However, it is argued that output measures should focus on the estimates of the volume of production of these services rather than their social valuation. As such, unit costs should be used as weights when aggregating outputs rather than marginal social valuations, which introduce a degree of subjectivity.

RÉSUMÉ

Ce compte rendu critique évalue le rapport de l'OCDE *Towards Measuring the Volume Output of Education and Health Services : A Handbook*. Par le passé, les coûts des intrants de l'éducation et des services de santé étaient utilisés pour les estimations de la production de ces industries. Le *Handbook* présente une analyse détaillée et des recommandations sur les données et les méthodologies appropriées pour améliorer les mesures de la production réelle pour l'éducation et les services de santé. Les tentatives de l'OCDE dans ce domaine sont très bienvenues. On soutient toutefois que les mesures des extrants devraient mettre l'accent sur l'exactitude de la mesure de la production de ces services dans les comptes nationaux plutôt que sur leur valeur sociale. Les coûts unitaires, plutôt que des valeurs sociales marginales, qui introduisent un certain degré de subjectivité, devraient par conséquent être utilisés comme poids pour l'agrégation des résultats.

NATIONAL ACCOUNTS HAVE BECOME a central source for evaluating living standards. Headline numbers of GDP are compared, policy makers worry about trends, and researchers delve into

details. However, a rather large deficiency exists in national accounts coverage of industries that are becoming increasingly important in advanced economies: health and education services. It is this

¹ The author is Chief, Policy Analysis Section in the Economic Studies and Policy Analysis Division of Finance Canada. I would like to thank Sir Tony Atkinson, Barbara Fraumeni and Andrew Sharpe for very helpful comments. The opinions expressed in this study are those of the author and do not reflect in any way those of Finance Canada. All the other usual caveats apply. Email: abIorwerth.Aled@fin.gc.ca.

gap that Paul Schreyer, Deputy Chief Statistician at the OECD, seeks to fill in *Towards Measuring the Volume Output of Education and Health Services: A Handbook* (Schreyer, 2010).²

Researchers have expressed concern regarding indicators of output volumes of many service sectors in the national accounts. But this shortcoming is particularly severe for education and health services, which are largely provided by government, since there are limited market output values or prices. Even if some share of output is provided by the private sector, the partial coverage by insurance or significant government involvement in price setting will mean that prices are not meaningful indicators of consumer choice or marginal valuation. As *The Economist* (2012) noted, “Americans spent \$2.6 trillion on health-care in 2010 ... yet few of them have the faintest idea what any treatment costs or how it compares with any other treatment.” Although prices for attending private-sector firms are available, one might be leery of applying these prices to their public-sector equivalents.

As a result of these problems, the national accounts traditionally incorporate these services at input cost: the value of output of education is equal to the value of all the inputs used in its production. In turn, this output=input treatment implies that the national accounts have nothing interesting to say about their productivity: it is always unitary and its growth is zero. This treatment can distort the national accounts, as the measure of aggregate real output of education and health services is not par-

ticularly meaningful. Given that many of these services are provided by government, this treatment may even contribute to popular depiction of government as a consumer of resources rather than a producer of services! Indeed, I was surprised to learn in reading the *Handbook* that during the formative years of national accounting, some did not feel that government services contributed to GDP.³

Over recent years, statistical agencies have started to move to rectify this situation by developing more sophisticated measures of government output. Professor Anthony Atkinson reviewed many new practices and offered further guidance at the behest of the Office of National Statistics (ONS) in the United Kingdom (Atkinson, 2005).⁴ Paul Schreyer’s task in this *Handbook* is the more daunting prospect of going from the ideal approach to the practical, and delivering concrete suggestions for output measures covering health and education services for inclusion in the national accounts, and this on a comparable international basis.

The *Handbook* is the latest instalment of the OECD’s efforts to bring some degree of coordination and comparability in data design and statistical methodologies between member states for education and health services.⁵ Of course, in addition to basing recommendations on statistical or economic theory, the advice of the OECD on suitable output measures will be taken bearing in mind whether international data will be available. Although trends over time in output for one country are informative by themselves,

2 The handbook, released by the OECD as Statistics Directorate Working Paper 2010/02, can be accessed at http://www.oecd.org/LongAbstract/0,3425,en_2649_33715_45115957_1_1_1_1,00.html

3 Non-market production has been fully recognized as contributing to GDP since the 1968 System of National Accounts (Schreyer, 2010).

4 For a discussion of the *Atkinson Review*, see ab Iowerth (2006).

5 This *Handbook* is an update of OECD (2007). *Eurostat* also plays a driving role, probably because budget transfers in Europe are linked to nominal GDP per capita and hence a consistent basis for defining GDP is required. Steps in North America are more tentative. Gu and Wong (2010) experiment with different approaches in education for Canada. The U.S. Bureau of Economic Analysis has conducted research in a number of areas but they judge their research to be preliminary and recommend further research (BEA 2011:9-4). Further analysis for the United States is conducted in Abraham and Mackie (2005). The Australian experience is discussed in Trewin (2004).

there is scope for significant insight if international comparisons of both levels and growth could be made.

There are several practical challenges in moving to output measures for health and education services that are independent of input measures. First is the obvious challenge of developing or obtaining data on output. A second difficulty is incorporating the impact of quality change. Finally, the appropriate weighing for aggregating different types of education and health services is unclear. There are related methodological problems. Clearly much work has been done by the OECD, particularly for education, but challenges remain, notably in measuring quality of health-care provision. It is therefore likely that the *Handbook* will remain a work in progress, and be regularly updated as new methods and data are developed. This conclusion is clear to the OECD as well, given their use of the word “towards” in the title of this volume.

The *Handbook* is mostly aimed at practitioners in the field. A certain amount of background in national accounting is needed to understand the various concepts. Given the very broad area that the *Handbook* covers, some additional illustrative real-world examples of various issues for a general reader may have been suitable. Indeed, for users of the national accounts such as myself, a short companion piece by the OECD to explain more succinctly the advantages and appropriate use of the data may be useful at some point. In this review article, I propose to spend most of my time in explaining some of the key concepts for a more general audience.

I welcome the efforts undertaken in developing this *Handbook*. However, I would have liked greater clarity in the *Handbook* when discussing production of health and education services. These are produced services, so their estimation in the national accounts should be akin to production of other services. There are no prices for the output of health and education and con-

sequently researchers have undertaken considerable effort to develop marginal social valuations of these services. These valuations could be used to aggregate and value outputs, a stance which the *Handbook* seems to support in principle.

But this seems to me to step beyond the role of the national accounts in valuing production. At this time, given the enormous difficulties in developing adequate and objective valuation metrics, it may be a step too far to incorporate these valuations in the national accounts. In this respect, I echo the view of Murray (2005) and Heikkinen and Hautakangas (2008) that the national accounts should focus on production rather than trying to incorporate social valuations. In the end, the *Handbook* favours using unit costs because of practical concerns related to developing adequate marginal valuations. But, for clarity, perhaps the *Handbook* should have been restricted to the valuation of production. Indeed, making contributions to improved measurement of output volumes is a major contribution in itself.

In the first main section of the article, I discuss why improving the national accounts, and hence improving volume output measures of education and health services, is desirable. In section two, I outline some of the key concepts in the area of measuring non-market output quantities and values. Section three provides a more detailed review of the *Handbook*. Section four concludes.

The Desirability of Improving the National Accounts

The national accounts play several crucial roles in understanding the state of the economy. They provide estimates of the value of economic activity, show the allocation of resources by industry, expenditure category, and income group, and allow for an assessment of trends in living standards. This information can be used

to identify areas where performance can be improved.

Although many metrics may exist to determine the performance of individual workers or institutions such as hospitals or schools, there is value in developing national accounts-compatible measures for the education or health system as whole. Such measures would allow better monitoring and analysis of economy-wide resource allocation issues. There is concern, for example, regarding the growing share of resources being allocated to healthcare (Hall and Jones, 2007). However, debate on whether this is inappropriate cannot advance unless there is evaluation of how effectively these resources are being used, and whether they lead to commensurate increases in output and welfare.

The contribution that national accounts-based measures could make is illustrated by preliminary research cited by the *Handbook* in Deveci *et al.* (2008). They examine output measures for Danish healthcare, and find that the input-price index increases by much more than the output-price index, independently of output. Using the output price index to deflate total expenditures would lead to a greater growth in healthcare volume output than if input prices were used. Such findings buttress arguments that increased spending on healthcare is productive (or of course the reverse if opposite results were found).

To the extent that education and health services are provided by government, a second role is improving government productivity. With increasing pressure in these times of austerity to ensure that government is efficient, the absence of meaningful, reliable, and accurate output measures means that incorrect assumptions may be made on the opportunity cost of cutting spending, or on drawing useful lessons on how government productivity can be improved. This is not a new challenge. Responding to spending increases in earlier times, economists writing on

the importance of measuring government output asked: "Has the government's productivity declined so that part of the rise in input reflects the need for more resources in order to maintain a given volume of services? Or has productivity risen, thereby causing government's output to rise even more rapidly than its input?" (Fabricant and Lipsey, 1952).

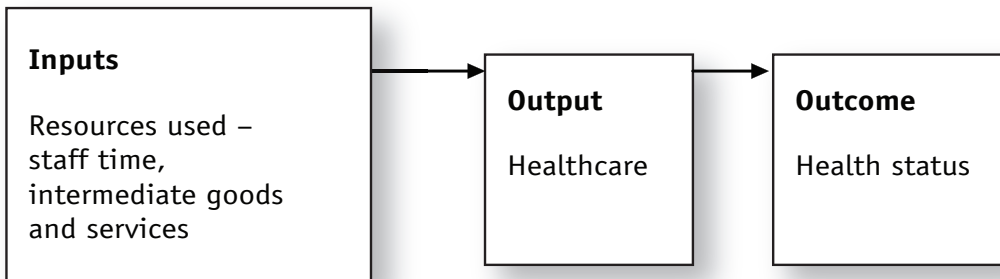
Better volume output measures for health and education services may also give governments pause before spending additional resources since these resources may not result in significantly increased output. Increasing output may involve difficult institutional reform, not just spending additional resources, so developing output measures can therefore encourage such fundamental reform. Ensuring that resources are spent effectively is difficult when there is a paucity of appropriate data; better data allow better questions to be asked. The Ontario Government recently established a commission headed by Don Drummond to examine fiscal challenges in the province. Their report (Commission, 2012:13) signaled a clear and important role for better information on the output and outcomes of government activities:

The best public service would set clear objectives, use proper metrics to measure progress and provide clear accountability for those expected to meet the objectives. It would benchmark itself against the best in the world. It would constantly evaluate priorities; if a new priority is identified, others would move down a spot and some, now outdated, would be discarded. It would drive relentlessly towards effectiveness and efficiency.

It should be borne in mind, however, that data gathered as metrics to improve performance within particular institutions may not be appropriate for the national accounts. For example, as discussed in Atkinson (2005), performance indicators need to be simple and precise, but not necessarily consistent over

Chart 1

The Definition of Inputs, Output, and Outcome: An Example from Healthcare



time. On the other hand, data produced for the national accounts can be transformed in complex ways, but do need to be consistent over time.

Finally, GDP, for better or for worse, has become an indicator of living standards. But the current treatment of non-market services means that information contained in the national accounts is inadequate for an evaluation of living standards and their growth, especially across countries with different national conventions for measuring real government output and with different shares of non-market services in GDP.⁶ Analysis must differentiate between results based on the business sector or on a total-economy basis. The current shortcomings of measuring non-market output are reflected when knowledgeable researchers exclude this sector from their calculations of productivity growth. Indeed, in its assessment of the adequacy of measures of economic performance the Stiglitz Commission argued that it is important to come to grips with measuring non-market output in order to have a satisfactory measure of economic

performance and living standards (Stiglitz *et al.*, 2009).

Economic Concepts, Theory and the National Accounts

Outputs versus Outcomes

Before embarking too far in this review, definitions of inputs, output, and outcomes are needed (Chart 1). As with all goods and services, production of education and health services involves the transformation of inputs into outputs. Inputs issued in the production of education and health services have the same purpose as for market-provided goods and services. For example, inputs would include the hours worked by teachers, the capital consumption from the capital goods such as computers, and intermediate inputs such as teaching supplies as intermediate goods. The time of healthcare staff and intermediate inputs such as pharmaceuticals are the inputs into hospital production. It is these costs that are currently—and inappropriately—recorded as the “output” of non-market services under the traditional method of national accounting.

⁶ For 2008, the OECD reports that 6.1 per cent of the OECD’s GDP is spent on education with several countries spending more than 7 per cent (OECD, 2011a). Just less than 9 per cent of GDP was spent on health with the United States at over 16 per cent of GDP (OECD, 2011b). As discussed in Box 1, the *Handbook* undertakes a preliminary analysis of incorporating secondary education outputs into the national accounts. This would, for example, raise Germany’s GDP per capita from 5 per cent above the OECD average using the current approach to 7 per cent. In contrast, Sweden’s GDP per capita would be reduced from 10 per cent above the OECD average to 8 per cent.

A challenge in measuring the non-market sector is the distinction between what is produced—which is measured by “output”—and what society values—which is measured by “outcome”. Normally, one might think of a widget as the output produced by an industry, and consumers gain utility from consuming that number of widgets, so output and outcome are the same.

The situation with non-market goods and services is more complicated:

- There is no market for the output and therefore it is hard to identify and agree upon what constitutes output. Many indicators and proxies exist, but they need to be sorted into what are the most appropriate indicators for either output or outcome. This sorting out requires understanding the structure of the sector in question; and
- The output from institutions supplying non-market services contributes only a part of the outcome. The individual efforts of students also contribute to the education outcome as well as the output of education institutions.

In turn, the distinction between outputs and outcomes is important when interpreting data. Are the data to record production of a good or service independent of outcomes, or are they to directly measure “social welfare” or well-being? I would argue that it is more important for the national accounts to accurately capture the output of health and education services in terms of production rather than the value society places on that output. First, having accurate measures of production is important in its own right and, second, the social value of education and health services is too subjective to be put in the national accounts.

Smith and Street (2007) report that there was no consensus among participants at a workshop on the measurement of education and health services about what constitutes the primary purpose of education. For education as a whole, one

may think that the outcome is an “educated person”, but one can well imagine the philosophical quagmire as one tries to place an economic value on such an outcome, or even agree on its quantification. The education system’s output contributes to the desirable outcome—whatever that is—and it is that output from production that the national accounts should try to capture. Furthermore, the final outcome often comes from both the production of either health or education services on the one hand, and individual effort on the other. For example, education outcomes come from the output of the education system—teachers in front of a class—and the endeavour of pupils. This joint production means that outcomes cannot be used to give information on the structure of production.

One of the key tasks in the *Handbook* therefore is to provide clear guidance on what constitutes appropriate measures of outputs within the context of the national accounts, and it does this. For example, test scores are results of education, but they are the joint result of education and individual effort. As such, the *Handbook* argues they are an inappropriate measure of education output but they could instead be appropriate proxies for outcomes. What appropriate means in this area may be a subject of debate and the *Handbook* goes through its justifications of why certain measures are adopted and why some are rejected in considerable detail.

Implications of outputs differing from outcomes

There are additional challenges the *Handbook* must tackle beyond recommending particular output measures. In the simple case of market-provided goods, the statistician obtains a series of prices and quantities. From these two series, the usual formulas of index theory can be used to derive measures of real output and productivity growth. However, neither of these series is normally available for non-market goods and ser-

vices. The *Atkinson Review* made substantial contributions in showing the importance of obtaining reliable measures of output quantity for non-market services, but delved less deeply into certain intricacies. Two challenges in particular remain in this regard. The first and more basic challenge for non-market goods and services is that there are no price indexes. The second challenge—common to market-provided goods as well—is to account for quality change.

Weights

The absence of prices indexes is a problem because aggregation of quantities is normally done using prices as weights. Prices reflect the marginal utility on social value a consumer places on a product and therefore act as a suitable weight in aggregating. Consequently, even if satisfactory volume indexes for different types of non-market outputs could be compiled, the absence of prices means that the challenge of aggregating these volume indexes remains. Can appropriate weights be found to substitute for prices?

Diewert (2010) examines the implications of using different measures for valuing and aggregating quantities, given that some information may not be available to the statistician. His suggested hierarchy of appropriate weights for valuing non-market outputs is:

1. Market prices or purchaser's valuation (first best);
2. Producers' unit costs of production (second best); and

3. An output-price growth rate equal to an index of input-price growth (third best).

The third of these is the practice currently employed in many countries when there are no data available on volume outputs. In other words, the only proxy we have for changes in output prices is changes in input prices. But in many cases, some data exist on prices, so the question is whether there are enough data to adopt the first- or second-best approach.⁷

If some market price data are available on market goods comparable to non-market goods (e.g. education provided by private schools), hedonic methods could be used to estimate prices. An interesting approach discussed by Diewert is to introduce output measures of non-market services as inputs into sectors of the market economy within an estimated general equilibrium model. Although requiring too much information to be practicable at the moment, this approach would in theory allow implicit prices for non-market services to be derived.⁸

Given the absence of either direct price data or of data for comparable market-provided goods and services, one is led to a dilemma when proposing weights for output volumes of health and education. Diewert notes that “For many purposes (including the measurement of welfare), the desired conceptual price for each type of medical service is a *household marginal valuation price* or a *final demand price*, i.e. the price that a household would be willing to pay for an extra unit of the service. But it is difficult for experts to agree on what the appropriate final demand prices should be. This lack of

7 Diewert (2008) examines the implications of different approaches in the absence of comparable market prices and in the presence of quality change. He demonstrates that using costs as weights can generate measures of productivity growth that are informative—relative to the use of only input cost data—if data are available to calculate input requirement per unit output. However, these indexes may be misleading in the presence of technological change.

8 A similar notion is used in Wang *et al.* (2009) for banking. They argue there is no clear basis for measuring the output of banks, as they often do not charge explicit fees for services but instead make money from the spread between interest rates charged and paid. To incorporate the systematic risk of bank loans, the authors employ a dynamic, stochastic, general equilibrium model to clearly derive and define the marginal unit of bank output.

expert consensus puts statistical agencies in a difficult position since their estimates of output and inputs should be *objective* and *reproducible*.” (Diewert, 2008:4). These arguments would seem to militate against using marginal valuations that are surrounded by controversy.

The *Handbook* reluctantly concludes unit costs— Diewert’s second-best approach— should be used as weights on practical grounds, although the *Handbook* is also sympathetic for some sort of marginal valuation by consumers. Other arguments advanced in the *Handbook* to support using costs include:

- Even if prices were observable, they are heavily distorted in this area by the presence of information problems, monopolistic suppliers and—in the case of health services—that observed prices would be based on insurance valuations rather than willingness to pay; and
- In a democracy, the valuations placed on services could reflect social preferences for them. As such costs are reasonable proxies for willingness to pay.

There is some equivocation in the *Handbook* on appropriate weights for aggregation, a discussion that merits greater clarity. The nature of weights is not a major topic with market-provided goods and services. For such goods and services, the marginal valuation placed by society is the price paid for it and, if we further assume perfect competition, then that price equals the marginal cost. As a result, the distinction between output and outcome is not normally a conceptual barrier because marginal social value equals marginal cost. Of course, there remain issues of imperfect competition or externalities driving a wedge between society’s valuation and the cost of production. However, for non-market goods and services where outcomes are not the same

as outputs, I would argue that costs would be more closely linked to output whereas marginal social valuations are linked to outcomes.

It would seem to me that the national accounts should reflect the value of production, and this would argue in favour of using costs as weights. This approach has several additional merits:

- Overall consistency in the national accounts. There is no convention in the national accounts for looking at externalities—pecuniary or non-pecuniary. Goods are included at the price paid, not at prices consumers should pay for them, even if they do have spillovers. As mentioned in the *Handbook*, the benefit of installing air bags in cars is not reflected as a value to consumers beyond additional costs reflected in auto-sector output; and
- Introducing a separate measure of willingness to pay may end up with the value of output not equal to the value of inputs, and the need for the resultant surplus or deficit to be reflected as some sort of profit or transfer elsewhere in the accounts.

Quality

Incorporating quality changes is a general challenge in the national accounts. Hedonic approaches have gathered interest in recent times. But with the general absence of data, the problem of incorporating quality may be even greater in the non-market sector. The problem is particularly acute in health services as new treatments and drugs become available.

Stratification of the data is an implicit way of controlling for quality: different qualities of a product will be treated as different products.⁹ The quality of a secondary education differs from a post-secondary education, so stratification reflects the fact that different qualities of service will be provided at different costs and

⁹ Aizcorbe *et al.* (2003) examines the finer disaggregation as an alternative to hedonic methods in controlling for quality differences for microprocessors. They quote Zvi Griliches saying that ‘If you have the right kind of data, you don’t need hedonics.’

that outputs cannot be simply added up: adding up the number of pupils in secondary schools and the number of students in university and calling the sum the output of the education sector is not suitable. Finer disaggregation will also allow resource reallocation to be detected.

Direct methods of incorporating quality may also need to be adopted through adjusting volumes or weights. In this case, the *Handbook* argues that outcome measures should be used to adjust outputs so that quality may be taken into account. Here again, however, one runs into the issue of subjectivity, as there may be disagreement on which outcomes are more important.

Although there is extensive discussion of potential quality metrics in the *Handbook*, there is a reluctance to strongly advocate for particular measures. This is probably the appropriate approach. The *Atkinson Review* recommended that while priority should be given to work on quality-adjusting outputs, “a relatively high threshold should be set for their introduction into the National Accounts; in particular, ONS should not introduce quality adjustments until it is assured that the dimensions covered are sufficiently representative.” (Atkinson, 2005:191).

As a side note, there is no discussion in the *Handbook* of using some of the quality measures to correct for the quality of inputs; indeed, some of the quality measures discussed may be more appropriate for use in this manner.

The Handbook: An Overview

Turning now to the contents of the *Handbook*, its first chapter outlines conceptual issues, which forms the basis for the analysis and guidance in the chapters that follow. A major contribution of the *Handbook* is the clear link between the theory and the guidance on appropriate output measures. The second and third chapters

analyze the appropriate indicators for the output of the education sector, the second chapter over time and the third across countries. Chapters four and five repeat the pattern for health services. There is much discussion of the economic concepts and how they relate to the principles of national accounting. Tables are included for cross-country comparisons of current practices. A second major contribution of the *Handbook* is documenting the large amounts of input received not only from national accountants, but also from specialists in education and health. There is extensive discussion of different measures and methodologies across countries.¹⁰

As mentioned, international comparisons will be informative in these critical non-market services because of the different institutions and policies followed across countries. In this respect, comparison of the output levels will be required, which necessitates some form of purchasing-power parity adjustment for education and health services. Making strides in this direction provides a clear justification for the OECD to help coordinate development of education and health indicators. They make some progress in this regard in the *Handbook*.

The *Handbook* lays out a clear path on developing measures of education and health services output, and it is worthwhile to outline that path before looking at the details.

The areas of health and education are too large to be useful units. Consequently, the first step is to stratify both areas into units of “production”: primary schools, secondary schools, etc. As well as grouping production units so that their production functions are roughly similar, stratification makes it easier for outputs to be grouped according to their quality. In turn, it would also allow technological change in the

¹⁰ The *Handbook* focuses on education and health services and does not discuss public administration. Yet the problems inherent in measuring volume output in public administration are similar to those for education and health services as the output is also non-marketed.

provision of a service to be detected and accounted for.

Given these units of production, the second step is to develop volume measures for each type of output. At this stage the *Handbook* provides useful rationales, based on both economics and data availability, about what would constitute satisfactory output measures.

The third step is to explicitly incorporate quality of output into the analysis beyond what has already been implicitly factored in through the stratification procedure in step one. Direct measures of quality can be incorporated such as test scores in education.

The fourth step is to develop methods to enable international comparisons of outputs. This stage involves development of some metric that allows upward or downward revision of volumes—analogue to purchasing-power parity—so that output is based on a common metric across countries.

Although the path above is clear, difficult conceptual issues and issues of data availability make its application difficult, particularly for quality measures. As the Stiglitz Commission noted, “While there are methodological disagreements about how to make the adjustments to quality or how to go about measuring output, there is a broad consensus that adjustments should be made, and even about the principles which should guide such adjustments. The disagreements arise in the practical implementation of these principals.” (Stiglitz *et al.*, 2009:12).

The *Handbook* makes progress along the path outlined above for education and health. But data problems mean that completely satisfactory results are not available at this time. In particular, although there is extensive discussion of the quality of healthcare output, the *Handbook* does not put forward proposals for explicit quality adjustment.

The challenges involved in looking for metrics in health and education differ. There is far

more heterogeneity in output of health services, since individual treatments differ in all sorts of minor ways. On the other hand, the outcome—health status—is relatively clear. By contrast, the outputs of education are relatively clear—such as the number of students—but what should be the desired outcome is less so.

Measuring Education Output

The *Handbook* starts by adopting UNESCO’s definition of education as the “organized communication of knowledge” as the basis for finding output measures. Stratification of education production follows the conventional separation of education leading to nine levels in total, such as by pre-primary, primary, secondary and post-secondary education. Although this seems like a detailed breakdown, my experience is that comparing post-secondary institutions across countries (or even across provinces in Canada) is quite complicated because of the mix of vocational and academic programs at some institutions.

The volume of output for these institutions is suggested to be the number of pupil-hours for primary and secondary education to reflect the provision of service to the pupil. The output is less clear for tertiary education since a much larger part of the education process is work by the student rather than the education institution. Consequently, the *Handbook* does not explicitly recommend a measure.

The *Handbook* provides detailed discussion of the various options to introduce quality measures for education, including on the basis of school inspections, class size, and standardized exam scores. The pros and cons of each are examined. Although imperfect, exam scores are preferred, as school inspections are subjective and not easily translated into quantities, and class size probably has a non-linear link to quality. The challenges of developing these indicators, which the *Handbook* wades through, are

perhaps best captured by the following summary of how to improve education outcomes:

“First, many of the traditional policies of simply providing more funds for schools or of not adding specific resources such as smaller classes do not provide much hope for significant improvement in student achievement. Second, a growing body of research shows that teacher quality is a primary driver of student achievement but that differences in quality are not closely related to teacher education and experience.” (Hanushek and Woessmann, 2011:479).

An alternative approach to evaluate the quality of education output is to use subsequent earnings performance by students. However, this human-capital approach suggested by Jorgenson and Fraumeni (1992) has the drawback of reflecting not only the output of the education system, but also the effort of the individual student. It is thus rejected.¹¹ Perhaps a closer approximation to the value provided by an educational institution could be based on Ferrer and Riddell (2002) who look at the incremental value that credentials bring after controlling for years of education, although this value may also reflect the ability or dedication of the student to finish a program.

The third chapter examines how the approach outlined above can be implemented for education services in practice in international comparisons. The output measures chosen in the previous chapter play an important role in determining how informative the international comparisons are. There may be significant differences in pupil-hours across countries—the recommended output measure—depending on school hours and the length of holidays, which translates into larger differences in output across countries than if a simpler measure of number of pupils were used. Since a relatively simple measure of output is chosen, the third

chapter is brief because these data are relatively readily available.

Quality differences across countries are likely to be greater than quality differences across time within a single country. Consequently, controlling for quality is critical in comparing country performance. This point is reinforced by Ferrer *et al.* (2006), who look at immigrant education and literacy in Canada. Among the factors that explain lower immigrant incomes in Canada, relative to the Canadian-born with the same education levels, is that the literacy skills of immigrants (as measured by the International Adult Literacy Survey) are lower than their foreign education might suggest. In other words, “differences in return to foreign versus Canadian acquired university education are entirely explained by foreign universities generating lower levels of (Canadian usable) literacy.”

The *Handbook* moves rather quickly to the different international tests available to control for quality. The Programme for International Student Assessment (PISA) administers a test for between 4,500 and 100,000 pupils in each country for mathematics, reading and science. Alternative tests are the Progress in International Reading Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS). The PISA test is preferred for quality adjustment in the *Handbook* on the basis that it covers a number of academic subjects, and is administered to 15-year-olds, which is more representative of secondary school populations. A shortcoming for time series analysis might be that the PISA test is only conducted every three years. At the moment there is no comparable test that could be used to compare quality of outcomes for tertiary institutions.

As mentioned previously, it may be difficult to determine what constitutes a desirable outcome of education and consequently, what quality

11 For discussion of these issues and the BEA experiments in measuring education output and outcomes, see Fraumeni *et al.* (2009).

Box 1 – An Example of Measuring Education Output

The methodology outlined in the *Handbook* is applied to expenditures per capita on secondary schooling across OECD countries in 2005. The traditional method of comparing inputs across OECD countries shows considerable variation in final expenditures per head on education. Countries such as Iceland, Israel and Australia put around twice as much input into education compared to Germany or Japan on a per capita basis. I take this to mean that the former countries have a combination of additional teachers and additional school supplies.

Using output volumes narrows these large cross-country differences. Those countries that had high inputs now have output that is only 50 per cent above the OECD average, which the *Handbook* argues is more plausible (without any reasoning given).

The ranking of some countries changes significantly depending on whether input or output measures are used: although Iceland ranks first whether input or output is used, Canada has the seventh highest input but only the 27th (out of 33 countries) highest output. Indeed, changing from input to output metrics lowers the implied contribution to education significantly for countries that spend heavily on education, such as Iceland, Australia, Sweden, Canada and Luxembourg.

Incorporating quality changes these rankings little. Quality adjusting this output raises Canada's ranking to 22nd. However, this ranking seems sensitive to the output measure chosen, as Canada ranks 29th in terms of the number of students as a share of its population (at around 19 per cent of the population). The highest ranked countries on the output measure—Iceland, Israel and Mexico—have around 30 per cent of their population in school. Since the output measure now simply reflects production, which is driven by the demographic structure of the population, there is not that much meaning to the ranking.

It would be interesting to examine additional metrics such as output per unit input, particularly over time. Now that the OECD is developing these output measures, the interpretation and data provided may change as well.

indicators should measure. One could well imagine that attaining a certain level of test results may be considered as short-sighted by those who value the education system as producing well-rounded individuals who can participate in society (e.g. Ravitch, 2012). Indeed, test scores may give undue weight to “academic” outcomes rather than to ensuring that individuals find occupations more suitable to their capacities, which in turn may be more vocational.

The *Handbook* provides experimental calculations using the proposed volume measures and

quality adjustments (Box 1). Although the OECD cautions that the findings are preliminary, they show the potential for interesting results. The results merit greater discussion than the *Handbook* currently provides, including more discussion of how to interpret the data appropriately. It is interesting to learn that these output measures may be more reliable than the input data currently used, as there are quality problems when comparing wages and salaries across countries. Unfortunately, this point was not elaborated.

Measuring Health Output

The definition of healthcare that guides the analysis in the *Handbook* is the “treatment of a disease or medical services to prevent a disease”. Clearly, there is wide heterogeneity in healthcare output as production is customized by individual. Nevertheless, the *Handbook* can take advantage of considerable administrative data that are available. It is at this stage that the OECD’s consultations with specialists are shown to be valuable.

In implementing the definition of healthcare, the *Handbook* argues that the ideal output measure would be complete treatments of a disease. Looking at complete treatment means that the output measure should reflect the passage of an individual through different institutions providing healthcare services. For example, an output such as a complete hip replacement would include visiting a general practitioner, examination and operation by specialists at hospitals, and rehabilitation services. This description also shows how difficult implementing this measure would be since data systems do not currently track individuals across different institutions. This approach is also difficult to adopt in the case of chronic illnesses or in services such as caring for the elderly that have no clear end point on an annual basis.

As a result of challenges in looking at complete treatments, the *Handbook* recommends looking at episodes of treatment of particular diseases given by an institutional unit as the measure of health output. The common approach in medical institutions is to use Diagnosis Related Groups (DRGs). These are composite bundles of hospital services for which a hospital is, for example, given a single predetermined reimbursement from government or insurers. A typical DRG system consists of 500 to 1000 categories, each system

different by country. A lengthy list of different output measures adopted in different countries is provided in the *Handbook*, which suggests the richness of potential indicators but also the challenges in ensuring comparability on an international basis.

There are problems with existing DRG data. The *Handbook* notes that there are some inconsistencies in the data, whereas data for national-accounts purposes would have to be relatively uniform over time. Smith and Street (2007) note concerns that DRGs do not incorporate technology change rapidly into their classification system.

The *Handbook* also makes recommendations for other health services. The output measure of institutions linked to chronic care may be limited to number of days of care provided. Other output measures are necessarily straightforward, such as the number of visits as a measure of output for general practitioners.

The *Handbook* argues that quality changes in the output of medical services cannot be made without some reference to the effects of medical services on outcomes. The outcome of healthcare is a state that consumers value, such as health status.¹² However, many factors will contribute to this outcome, including socioeconomic factors such as general economic conditions and inequality and poverty; individual behaviour such as exercise and diet; and environmental factors, such as the extent of pollution.

Assessing quality in healthcare would seem to be exceptionally difficult. Process quality depends on whether the right choices are made in treating the patient according to professional standards. To the extent that that is true, compliance rates with established procedures could be used as indicators, but there is insufficient data

12 Others may wish to incorporate other factors in quality measures, such as waiting time or the freedom of choice to choose one’s own physician or treatment.

on best practices at the moment. Other quality measures are country or institution-specific. The *Handbook* concludes that it is too difficult to recommend an explicit quality adjustment for health volume output at this time.

The *Handbook* undertakes preliminary analysis of international comparisons of healthcare, which again has the potential to be valuable work. Much has been written about differences across countries in healthcare systems and the markedly different shares of their national incomes that countries spend on healthcare. Nevertheless, international comparisons are currently made solely on total expenditures, or inputs. Developing output measures would necessitate a form of PPP conversion to ensure that comparisons of appropriate quantities are made. Spending more money as a share of GDP may be perfectly acceptable if more services are obtained, but less so if there is no commensurate increase in output.

One can well imagine the practical difficulties in data collection to shed light at these critical issues. The variety of healthcare outputs makes it difficult to identify a basket of health goods and services that can be compared across countries. There are differences in data gathering and definitions. Furthermore, institutional differences across countries may mean that there are a variety of incentives to supply varying quantities and qualities health services at different costs (and prices). In other words, the basket of healthcare goods and services might be endogenous to the institutional structure.

Despite these challenges, the *Handbook* presses ahead and looks at international comparisons of hospital services. These services constitute a large share of total healthcare costs and data are more readily available. The DRG approach is used as the basic unit of analysis with detailed descriptions provided of ideal data and 44 proposed case types (e.g. treatment of heart failure, normal delivery of babies, pneumonia,

etc). Clearly, there has been much consultation on these recommendations as they are drawn from a variety of expert groups. Nevertheless, challenges are identified such as the absence of an internationally-comparable DRG system and incomplete coverage. Proposals are made to establish greater standardization in this area.

Concluding Remarks

The *Handbook* takes it as given that measuring output of education and health services is desirable. Having poor measures of output and productivity for such a large part of the economy will give misleading impressions of living standards, particularly as these services may become more important with an ageing population. Despite clear arguments in favour of developing output measures, it is fair to ask whether there are any drawbacks.

In principle, as mentioned in section two, a government's use of better measures of output should lead to better questions and decisions. However, in practice, it could also lead to poorer decision making as well. If exam scores were to be developed as the appropriate metric for quality improvement over time, then there could be an incentive to make exams easier. Ensuring international comparability of tests by using PISA would remain vital to avoid this possibility.

Incorporating volume output measures for education and healthcare services into the national accounts is a valuable initiative and will lead to better information on the aggregate allocation of resources. Such data will be particularly useful if they enable comparable international comparisons to be made so that lessons can be drawn on how to improve public services. This effort should also be extended to other non-market services, such as public administration. Clearly, the OECD can play a constructive role in ensuring progress on these fronts, although the usual plea with OECD doc-

uments for greater brevity and clarity also applies!

The *Handbook* provides a valuable summary of the state of play on measuring volume output of health and education services around the world. Although one may quibble with many details, it is clear that there is a large amount of painstaking work going on. Much work remains to be done on thorny methodological and data challenges.

However, I believe effort should be focused on adequately capturing the volume output of education and health services rather than their outcomes. As such, their estimation in the national accounts should reflect the cost of production rather than social valuations. Despite the desire to make the national accounts as comprehensive as possible, a degree of conservatism and prudence is necessary, and a modesty of expectations. Consequently, I would restrict attention to providing production metrics for health and education and therefore use costs as weights. As acknowledged by the *Handbook*, one has to be a bit wary of including quality measures.

Despite this caution, the importance of these industries and the wealth of data and analysis discussed in the *Handbook* suggest they should be a prime candidate for greater analysis in satellite accounts (Aizcorbe *et al.*, 2008).

References

- ab Iorwerth, Aled (2006) "How to Measure Government Productivity: A Review Article on 'Measurement of Government Output and Productivity for the National Accounts' (The Atkinson Report)," *International Productivity Monitor*, No. 13, Fall, pp. 57-74. <http://www.csls.ca/ipm/13/IPM-13-Iorwerth-e.pdf>
- Abraham Katharine G. and Christopher Mackie, eds. (2005) *Beyond the Market, Designing Nonmarket Accounts for the United States*, National Research Council of the National Academies (Washington, DC: the National Academies Press.)
- Aizcorbe, Ana, Carol Corrado and Mark Doms (2003) "When do Matched-Model and Hedonic Techniques Yield Similar Price Measures?" Federal Reserve Bank of San Francisco Working Paper 2003-14. <http://www.frbsf.org/publications/economics/papers/2003/wp03-14bk.pdf>
- Aizcorbe, Ana, Bonnie A. Retus, and Shelly Smith (2008) "Toward a Health Care Satellite Account," *Survey of Current Business*, May, pp. 24-30. http://www.bea.gov/scb/pdf/2008/05%20May/0508_healthcare.pdf
- Atkinson, Anthony (2005) *Atkinson Review: Final Report, Measurement of Government Output and Productivity for the National Accounts* (Norwich, U.K.: HMSO) available at <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/public-sector-methodology/atkinson-review/index.html>.
- Bureau of Economic Analysis (2011) "Chapter 9: Government Consumption Expenditures and Gross Investment: Updated November 2011," (Washington, D.C.: Bureau of Economic Analysis). Available at: <http://www.bea.gov/national/pdf/ch9%20govt%20ce&GI%20for%20posting.pdf>
- Commission on the Reform of Ontario's Public Services (2012) "Public Services for Ontarians: A Path to Sustainability and Excellence," (Toronto: Government of Ontario). www.fin.gov.on.ca/en/reformcommission/
- Deveci, Nursen, Heurlén, Kamilla and Sørensen, Henrik Sejerbo (2008) "Non-Market Health Care Service in Denmark – Empirical Studies of A, B and C Methods," paper presented at the 30th General Conference of the International Association for Research in Income and Wealth, Portoroz, Slovenia. <http://www.iariw.org/papers/2008/heurlen.pdf>
- Diewert, W. Erwin (2008) "The Measurement of Nonmarket Sector Outputs and Inputs Using Cost Weights," *mimeo*, University of British Columbia. Available at: http://www.nber.org/public_html/confer/2008/si2008/PRCR/diewert.pdf
- Diewert, W. Erwin (2010) "Measuring Productivity in the Public Sector: Some Conceptual Problems," *mimeo*, University of British Columbia. Available at: <http://faculty.arts.ubc.ca/ediewert/dp1004.pdf>
- The Economist* (2012) "Shopping around for Surgery: Companies Try to Make Health-care Costs Transparent," February 4.
- Fabricant, Solomon and Robert E. Lipsey (1952) "Productivity in Government and the Output of Government Services," in *The Trend of Government Activity in the United States Since 1900*, Solomon Fabricant and Robert E. Lipsey (eds.) (New York: National Bureau of Economic Research), pp. 84-111.
- Ferrer, Ana, David A. Green, and Craig W. Riddell (2002) "The Role of Credentials in the Canadian

- Labour Market,” *Canadian Journal of Economics* Vol. 35, No. 4, pp. 879-905.
- Ferrer, Ana, Green, David A. and Riddell, W. Craig (2006) “The Effect of Literacy on Immigrant Earning,” *Journal of Human Resources* Vol. 41, No. 2, March, pp. 380-410.
- Fraumeni, Barbara M., Marshall Resinsdorf, Marshall, Brooks B. Robinson, and Matthew P. Williams (2009) “Price and Real Output Measures for the Education Function of Government,” in *Price Index Concepts and Measurement*, W. Erwin Diewert, John S. Greenlees. and Charles R. Hulten (eds.) (Chicago: University of Chicago Press), pp. 373-403.
- Gu, Wulong and Ambrose Wong (2010) “Investment in Human Capital and the Output of the Education Sector in Canada,” paper presented at the 31st General Conference of the International Association for Research in Income and Wealth. St-Gallen, Switzerland. <http://www.iariw.org/papers/2010/poster1Gu.pdf>
- Hall, Robert E. and Charles I. Jones (2007) “The Value of Life and the Rise in Health Spending,” *Quarterly Journal of Economics*, Vol. 122, No. 1, February, pp. 39-72.
- Hanushek, Eric A. and Ludger Woessman (2011) “Educational Outcomes,” *Economic Policy*, July, pp. 427-491.
- Heikkinen, Jani and Sami Hautakangas (2008) “Outcome and National Accounts”, paper presented at the 30th General Conference of the International Association for Research in Income and Wealth, Portoroz, Slovenia. Available at: <http://www.iariw.org/papers/2008/heikkinen.pdf>
- Jorgenson, Dale W. and Barbara M. Fraumeni (1992) “The Output of the Education Sector,” in *Output Measures in the Service Sector*, Studies in Income and Wealth, Vol. 56, Z. Griliches, (ed.), (Chicago: University of Chicago Press), pp. 303-343.
- Murray, Richard (2005) “A Review of the Atkinson Review,” Swedish Network for European Studies in Economics and Business. Available at: <http://www.snee.org/filer/papers/329.pdf>
- OECD (2007) “Towards Measuring Education and Health Volume Output: An OECD Handbook,” Working Party on National Accounts (Paris: OECD).
- OECD (2011a) *Education at a Glance, 2011* (Paris: OECD).
- OECD (2011b) “OECD Health Data 2011 – Frequently Requested Data,” (Paris: OECD)
- Ravitch, Diane (2012) “Schools We Can Envy,” *New York Review of Books*, Vol. 59, No. 4, March, pp. 19-20.
- Schreyer, Paul (2010) “Towards Measuring the Volume Output of Education and Health Services: A Handbook,” OECD Statistics Directorate Working Paper 2010/02, (Paris: OECD) http://www.oecd.org/LongAbstract/0,3425,en_2649_33715_45115957_1_1_1_1,00.html
- Smith, Peter C. and Andrew Street (2007) “The Measurement of Non-Market Output in Education and Health,” *Economic and Labour Market Review*, Vol. 1, No. 6, June, pp. 46-52.
- Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Paris: Commission on the Measurement of Economic Performance and Social Progress) Available at: <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>.
- Trewin, Dennis (2004) “Measurement of Services: Recent ABS Experience,” *Statistical Journal of the United Nations*, Vol. 21, No. 3-4, pp. 277-290.
- Wang, J. Christina, Susanto Basu and John G. Fernald (2009) “A General-Equilibrium Asset-Pricing Approach to the Measurement of Nominal and Real Bank Output,” in *Price Index Concepts and Measurement*, W. Erwin Diewert, John S. Greenlees, and Charles R. Hulten, eds. (Chicago: University of Chicago Press), pp. 273-320.