

A Framework for the Accuracy Dimension of Data Quality for Price Statistics

Geoff Neideck
Australian Bureau of Statistics
September 2007

ABSTRACT

Given the importance of price indexes for monetary policy and other economic decision-making, and the complexity of price index construction and compilation, ensuring the quality of price index statistics is a major focus of the efforts of prices statisticians within national statistical offices. One of the key aspects of data quality is data accuracy. Within the context of the accuracy dimension of the ABS Data Quality Framework this paper sets out a framework for quality assuring price statistics. Discussion of the application of the framework highlights the key issues and risks that need to be addressed to ensure acceptable data quality for price indexes. An emphasis is given to managing the risks associated with changes that are made both with index updating (price sample changes, product reviews, index rebasing, etc.) and with changes to processes and introduction of new developments. The paper draws on some recent learnings from ABS's price index compilation experience.

INTRODUCTION

In June 2006, the ABS reissued the March Quarter 2006 releases of its International Trade Price Indexes and Producer Price Indexes. The reissue resulted from incorrect exchange rates being applied to convert foreign currency transactions into Australian dollars. The error stemmed from a change in the format of the spreadsheet containing exchange rates sourced from a data provider that went unnoticed by ABS index compilers.

Production of price indexes is a challenging task for National Statistical Offices. The conceptual basis for index construction is complex and undertaking pricing and index compilation often require high level problem solving and decision-making. How then does an NSO ensure the accuracy of its indexes and avoid errors such as the one described above occurring?

This paper examines the range of issues and circumstances in which errors are likely to be made and sets out a broad framework for mitigating the risk of errors occurring. It does this by extending the accuracy dimension of the existing ABS Quality Framework. There is also a discussion of the most common reasons why errors occur in index production and a set of principles is suggested that should be followed to help minimise the chances of significant errors occurring

The ABS Quality Manual defines accuracy of statistical information as “the degree to which the information correctly describes the phenomena it was designed to measure”¹. There are two key aspects to ensuring accuracy of statistical data. The first is ensuring that correct methodologies are applied and the second is ensuring that those methods are applied correctly. This paper has less to say on the most appropriate methodologies adopted for price indexes but rather focuses on the things that are important in making sure that those methodologies selected are correctly applied.

Similarly, the accuracy of statistical data can be thought of in terms of sampling error and non-sampling error, or more broadly error associated with the design of the statistical methods and error introduced from sources outside the statistical design. Again, this paper is less concerned about the statistical error associated with the design of price indexes and more concerned with the types of “non-design” errors that can creep into price indexes. Some of these types of errors can be considered beyond the control of NSOs (at least without reviewing methods and processes) such as population under/overcoverage, respondents misunderstanding of questions, etc. Of particular interest in the context of this paper is preventing those errors over which price collectors and index compilers have, or ought to have, some control.

SOME OTHER RECENT EXAMPLES OF PRICE INDEX ERRORS

Despite the best intentions and preparations of statistical offices errors in statistics do occur. This is equally true of price statistics as of other statistics. Some errors go unnoticed while others are captured in time before data are released. Unfortunately, some errors are detected after the release of price indexes and revisions or errata have to be produced.

¹ *ABS Quality Manual*, v2005/01

In addition to the incident described above, ABS recently revised its Price Index for 'Other Than House Building' for Perth. This occurred when it was discovered that this index, which had been ceased and then reinstated at the request of a particular client, had been incorrectly linked to the previous index. In effect, the link period value aggregates were price updated twice when the procedure was undertaken manually outside of the ABS prices computer system. Checks at the time on the veracity of the 'linked' index failed to detect the error.

Some recent experiences from other NSOs

Some other incidents of price indexes revisions by NSOs are noted below.

Switzerland A 2004 Ottawa Group meeting paper from the Swiss Federal Statistical Office, describes how in reweighting energy items in the Swiss CPI volume data were inadvertently used instead of expenditure estimates. ²

Canada In July 2006, Statistics Canada notified users of its traveller accommodation services price indexes of revisions due a change in the implementation of price index formula. This revision stemmed from the previous formula application which inadvertently price updated expenditure weights. ³

New Zealand In July 2007, Statistics New Zealand reissued a number of Producer Price Indexes due to incorrect calculation of energy price movements from the June 2006 quarter to the September 2006 quarter, which occurred while introducing a new methodology. ⁴

EXTENDING THE ACCURACY DIMENSION OF THE ABS QUALITY FRAMEWORK IN RELATION TO PRICE STATISTICS

Within the ABS Data Quality Framework there are six dimensions to quality:

- Relevance
- Timeliness
- Accuracy
- Coherence
- Interpretability
- Accessibility.

This framework draws on the one developed by Statistics Canada (Brackstone, 1999) and provides a basis for defining statistical data needs and for managing the risks associated with delivering quality assured statistical output.

In terms of the accuracy dimension of the framework, the ABS Quality Manual presents a range of quality characteristics and indicators for each of these dimensions covering all

² Boesch and Verrmeulen, 2004

³ *The Daily*, Statistics Canada, 27 July 2006 and correspondence with Statistics Canada

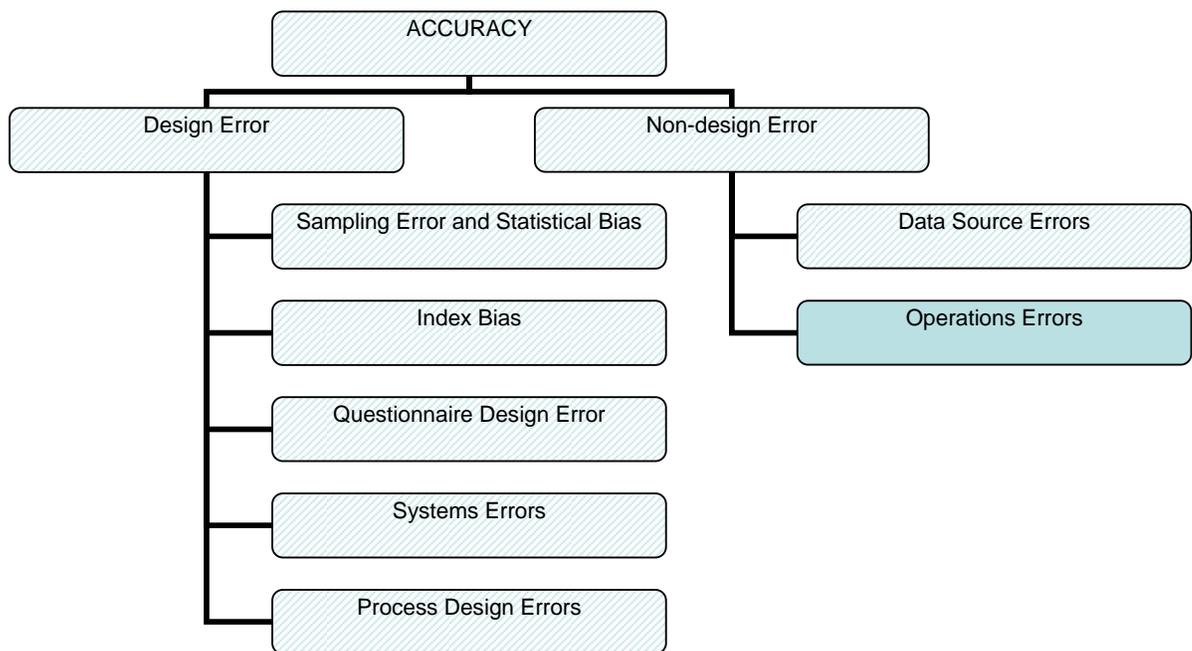
⁴ Statistics New Zealand, *Producer Price Indexes*, July 2007

aspects of the statistical cycle. It is fair to say that the ABS framework caters well for the statistical measures of data reliability (response rates, standard errors, etc.). Limited information is provided on the best approaches to managing “non-sampling” type error⁵.

The diagram below extends the ABS Quality Framework to cover the types of design and non-design errors that need to be avoided to ensure data accuracy. Design errors cover the sampling error, statistical biases and index biases that can occur with the design of price indexes. It also covers errors stemming from flaws in the design of data collection instruments (questionnaires, specifications provided to price collectors), and in the processes and systems used in price indexes. Design errors are not further discussed in this paper⁶. It should be noted that scientific sampling techniques are not used in ABS indexes except for its Labour Price Index.

Two types of non-design related errors are identified - those associated with source data errors and those arising from operations errors. Errors in source data may occur through respondents misinterpreting questions, accessing incorrect original records, failing to recall information correctly, and so on. Where data is provided from administrative sources inaccuracies can arise from changes to administrative procedures and wrongly recorded or incomplete information.

Operations errors are those errors introduced by the price collectors and index operations staff – the human errors that can occur. These are discussed in more detail later in the paper.



⁵ While this detail is not captured in the ABS Quality Manual, ABS has devoted much attention to these issues over recent years and encapsulated the issues and recommended responses in various ways e.g. in macro-editing strategies, the quality gates program, etc.

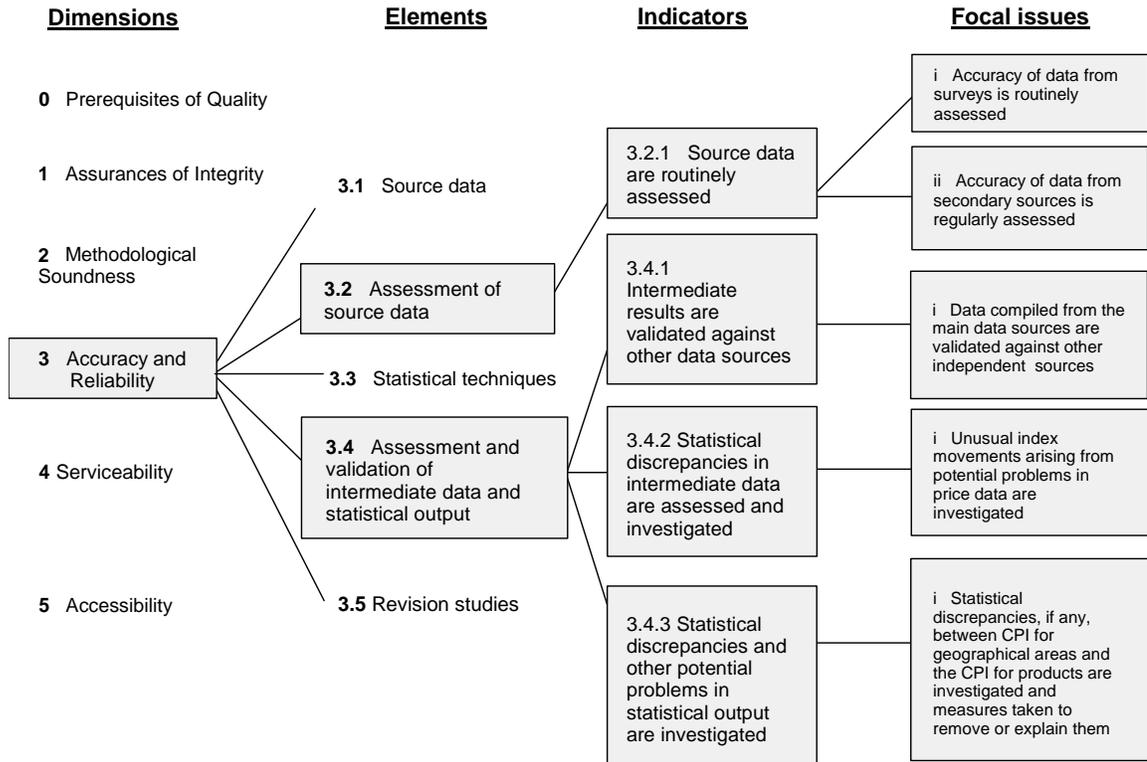
⁶ There is extensive literature on sampling error in general. In relation to sampling error in price indexes there is a good discussion in ch. 11 of the ILO Consumer Price Index Manual. The ILO CPI Manual also discussed biases in CPIs.

Comparisons with Other Frameworks

The framework above has similarities with the typology of errors set out in Chapter 11 of the ILO CPI Manual. Both frameworks cover the same range of errors but categorise them somewhat differently. The ILO Manual typology places more emphasis on sampling error. As mentioned above this is less relevant in the Australian context. Non-sampling error is also treated differently. In the framework presented here coverage bias (under and over) and non-response bias are covered along with sampling error as types of design errors. Both frameworks separately identify data source error ('response error' in the ILO manual) and operation error ('processing error' in the ILO manual).

The Statistics Department of the International Monetary Fund have produced Data Quality Assessment Frameworks (DQAFs) for both consumer price indexes and producer price indexes (IMF 2003). The following diagram presents the comparable elements of the IMF DQAF, in particular highlighting those aspects that relate most closely to eliminating 'non-design' errors.

**IMF Data Quality Assessment Framework
- Aspects relating to Operational Accuracy**



REASONS THAT OPERATIONS ERRORS OCCUR

There is a range of ways operations errors can occur in price indexes. The examples provided above describe instances of source data being misinterpreted or incorrectly used and index methodologies being inappropriately or incorrectly applied. These descriptions indicate how procedures went wrong. But why is it that they went wrong?

From investigations and reviews undertaken by ABS, both in relation to price indexes and to other statistical outputs, a number of critical factors have been identified that help in understanding why such errors occur. In understanding these issues it is then possible to identify the critical risk areas and develop strategies to manage those risks.

Critical risk factors associated with price index compilation are:

- Culture
- Change
- Education
- Documentation
- Engagement with stakeholders
- Spreadsheets and 'black boxes'

Culture

We can all recognise that the culture of an organisation, sometimes referred to “as the way we do things around here”, has an enormous influence on the quality of its services or outputs. Former Australian Statistician, Dennis Trewin, stated that quality is a fundamental aspect of an NSO’s operations and that the NSO:

“...depends on the actions of all its staff as all can have an impact on quality in one way or another. It cannot be left to a work group with designated responsibility for quality. Therefore, quality can only happen if there is a genuine quality culture within the organization.”⁷

Trewin set out many aspects that highlight the quality culture that exists in ABS and how these have been enhanced by particular quality initiatives. These initiatives continue. An overview of one of the latest initiatives on “Quality Gates” is shown in Appendix 1. However, maintaining a culture is challenging, especially where high turnover of staff is experienced.

As with any area of a statistical office, it is fundamental for a price index group to have “a genuine quality culture”. Indications that the challenge of establishing or maintaining such a culture is not being fully met can be seen when index staff:

- understand how to undertake processes and tasks but do not have an understanding of why those processes and tasks occur

⁷ Trewin 2002

- fail to take an ‘end-to-end’ view of the index compilation process i.e. do not appreciate how their work contributes and the links and dependencies that exist with other parts of the index
- assume that source data inputs are fully reliable and consistent with past inputs
- don’t take an inquisitive approach to follow-up on issues that they identify
- do not see data quality as part of their role.

Change

Occasions when changes occur to processes, methods, systems, etc. are times of biggest risk of errors. Quality assurance of data would be relatively easy if processes and methods did not change. However, price indexes are continually changing. Whether it is period to period changes in product and outlet samples, regular updates to weights and index structures or changes to introduce new methodologies, price indexes are frequently changing. Of course these changes are essential to maintain the relevance of indexes, to make necessary updates and index series links and to make improvements to methods.

Where changes are a regular part of the index processes e.g. products ceasing or new products being introduced into an ongoing sample, the risk of errors tend to be lower, as these are regular changes that staff are familiar with and standard editing approaches focus on.

Less frequent changes even if conducted on a regular basis may be more susceptible to error as they are often more complex and staff less familiar with the methods involved. Even more risky are the introduction of new methods. Some of the examples cited above fall into this category. Changes to new methodologies require staff to understand new things and often to think differently. Transfer of knowledge from the designers of the new methods to operation staff is critically important so they can understand both the ‘why and how’ of the new methods.

An added risk occurs when change is made outside of existing systems. This often occurs when established computer systems are not designed to cater to the new methods or time does not permit integration into the existing system. Ad hoc approaches are often used and standard validation and checking may not be appropriate or not applied as rigorously.

Education

To effectively undertake their work, price index staff need to be knowledgeable about price index concepts and practices. Knowledge of basic concepts and procedures can be acquired adequately in initial training and on the job experience. However, many aspects of index compilation are complex and much of the decision-making that is required is not straightforward.

Field collection officers may be required to make decisions about best sellers, product replacements and quality changes in products. Compiling staff need to make decisions about quality adjustment, product and outlet substitution, interpretation against independent information, and so on. In making these decisions staff need to have some

understanding of the methods on which they are making their decisions. For the most part these will involve standard price index methods. Relative straightforward procedures can be developed and staff can be readily trained to follow these procedures. From time to time however, staff will be called on to make decisions that involve more advanced understanding of index theory and practices. There are often choices to be made what practices to adopt, such as which index formula to apply or what quality adjustment method to use. Not all staff will need to be equipped with the necessary understanding to make these choices. However, it is imperative that these skills are developed and available within the index teams or elsewhere in the statistical office (e.g. within a methodology group).

Documentation

Lack of documentation or poor documentation makes it difficult for staff to undertake their work. This is particularly so for new staff. In surveys of ABS staff which ask questions about what would help them do their job better, invariably better documentation is high on the list of responses. It is not always true that better documentation will improve staff performance but staff cannot be expected to perform well if documentation is inadequate. Inadequate documentation results in staff 'filling' in the gaps as best they can or making assumptions about how best to undertake the task. Such ways of working can easily lead to errors.

Engagement with Stakeholders

In the context of index compilation work, the stakeholders in the processes are data suppliers, compilers of other parts of the index and managers. The reliability of data produced depends on a good understanding of data and information at every stage of the index compilation process. Failure to know about or understand key issues that arise, or changes that have been made in early stages of the process can undermine the quality of decisions that are made downstream.

Spreadsheets and 'Black Boxes'

Special mention should be made of two particular areas of operations where risks of error can be high.

As mentioned above work that is undertaken outside of existing systems and in ad hoc ways can often be susceptible to errors. One of the main ways that this occurs is through the use of spreadsheets. Spreadsheets are the cottage industries of statistical processes. Everyone has their own approach and practices vary widely. Documentation of methods, labelling and formatting is often poor, making checking of work difficult. Often the work goes unchecked. Even where the work is accurate it may become difficult to replicate when taken over by another staff member

Processes that occur that transform data where the workings of those processes are not transparent are often referred to as 'black boxes'. A staff member may see data going in and different data coming out but is not fully aware of the transformation that has taken place. Failure to understand the transformations that occur can lead to acceptance of incorrect results as there may not be well formed expectations of what answers are sensible. Similarly errors in input data may not be picked up if there is not a good understanding of the transformation process that has taken place.

KEY PRINCIPLES FOR MANAGING FOR DATA QUALITY

In considering where the risks to data quality are most likely to come from, it is possible to derive a set of key principles that will assist an agency in lessening the risks of errors occurring. These are:

1. *Developing and Sustaining a Quality Culture*

Whether it is explicitly stated or not, data quality is a core value of any statistical agency. Many agencies also see it as a corporate objective to deliver high quality outputs. All staff should see themselves as being contributors to the quality of those outputs. Staff in statistical (index) operations areas should be able to have a clear line of sight from their work to the outputs they are directly contributing to through to the corporate objectives they are helping to meet.

Information should be made available to staff on corporate objectives on data quality and quality initiatives that are being implemented. Training for staff should emphasise the importance of quality assurance and staff performance management should incorporate expectations on levels of quality to be attained.

2. *A program of staff capability development*

Clearly training is critical to developing competent price statisticians and practitioners. Formal courses, seminars, and on-the-job training are all important. Continuous learning approaches should be adopted. These can be best carried out in conjunction with staff development plans with coaching and mentoring programs.

Index program managers need to ensure that all the required skills, knowledge and capability are catered for across their teams (or available from support areas). Staff recruitment and development programs should be undertaken from a top down perspective to ensure these capabilities are established and maintained over time.

Part of education of price staff needs to be on understanding roles, responsibilities and accountabilities in relation to data quality.

3. *Managing Change*

Given the heightened risks of errors in price indexes at times when changes are being made, sound change management strategies should be put in place. Many change management models exist that provide a sound basis for working through change processes. Generally, the larger and more complex the change the greater the necessity to have in place formal change management plans. Formal plans should accompany any major review work, index redesign or computer system redevelopment. While full scale change

management plans are not necessary for more routine or smaller scale changes, the basic principles of change management should be adhered to - good project planning, communication with stakeholders, documentation of changes, training, etc.

In terms of minimizing the risk of errors, any changes should be accompanied by appropriate validation and checking. Checks should include not just confirmation that any calculations have been correctly made but validation of the conceptual basis for the change and of any assumptions that have been made. Assessment that implementation of new methods has been completed successfully should be made⁸.

4. *Good communication with stakeholders*

Many of the risks of errors can simply be eliminated by good communications. Regular meetings that focus on data expectations, issues arising in current processing and clearance of intermediate and final data can help ensure a full understanding of the quality of data.

5. *Appropriate documentation*

Documentation is a way of formally passing on information to assist staff to do their job. It ensures that processes that need to be followed are set out so that they can be completed fully and repeated in consistent ways, even when staff change. Good documentation should assist staff in understanding not just what to do in their job but also why tasks and processes are being undertaken. It should also help them understand where procedures need to be strictly followed and where they may be required to exercise their judgement.

6. *Evaluation and Review*

Following the introduction of any major change to price index procedures and methods it is good practice to undertake an evaluation. The focus of such an evaluation should be on whether the objectives of the new methods or procedures have been met and whether they are working efficiently and cost effectively. Similarly, for ongoing processes a regular program of reviews should be put in place to confirm whether statistical and organisational objectives are being met. Reviews should focus on high risk areas and where possible be undertaken independent reviewers.

⁸ Statistics Canada's Quality Assurance Framework has a good discussion on managing the implementation of new statistical methods.

Appendix: Quality Gates in the Index Compilation Process

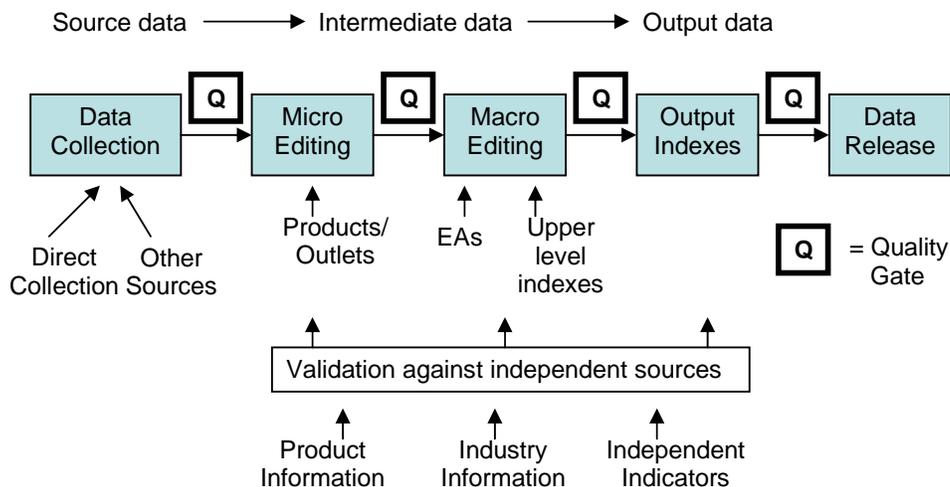
In order to ensure accuracy throughout the index compilation process, the concept of quality gates is useful in identifying key points at which it is important to assess data quality before data passes from one stage to the next. The ABS Quality Manual states that quality gates can be used “to improve the visibility of data quality in statistical processes, as well as being used to measure and monitor quality in real-time”.

In determining which process should be monitored index managers should base their decision on a top down assessment of the statistical risks associated with the particular index process. Processes with a higher risk of quality degradation and error should be monitored more intensely than those with lower risks. Business analysis or process mapping may be useful in determining which phases of the statistical process are most appropriate for quality gates.

The quality gate consists of a set of indicators that must be signed-off before data is accepted at the next stage of the process. Indicators may be quantitative or qualitative and need to be agreed by all stakeholders, that is those that are passing on the quality assured data and those who are receiving the data.

The diagram below illustrates the possible application of quality gates in a typical index compilation process.

Index Compilation Process and Quality Gates



REFERENCES

ABS, *Quality Manual*, v2005/01, Internal ABS Corporate Manuals

Allen, B. *Qualifying quality - issues of presentation and education*. Presented to the International Symposium on Data Quality, Quebec, Canada, 16-19 October 2001

Brackstone G., *Managing Data Quality in a Statistical Agency*, (1999) Survey Methodology, Vol. 25, no. 2, Statistics Canada

Boesch, R and Corrine Becker Vermeulen, *Quality assurance of the Swiss consumer price index*, Ottawa Group Meeting, Helsinki, Finland August 2004

International Labour Office, *Consumer Price Index Manual, Theory and Practice*, 2004

International Monetary Fund, *Data Quality Assessment Framework for the Consumer Price Index*, July 2003

International Monetary Fund, *Data Quality Assessment Framework for the Producer Price Index*, July 2003

Statistics Canada, *Statistics Canada's Quality Assurance Framework 2002* Catalogue no. 12-586-XIE

Trewin, D. *The Importance of a Quality Culture*, Survey Methodology, Vol. 28, No. 2, Statistics Canada, December 2002