QUALIFYING QUALITY - ISSUES OF PRESENTATION AND EDUCATION[1]

Bill Allen[2]

ABSTRACT

The ABS produces many statistics which are used by the government and wider community in making informed decisions. However, if these decisions are to be truly informed, it is essential that the users of these statistics are able to understand the limitations of the statistics and use the data in an appropriate context. As a result, the ABS has initiated a project 'Qualifying Quality', which focuses on two key directions; presentation and education. Presentation provides people with information on the quality of the data to help them answer the question "Are the data fit for purpose?", while education assists those people in appreciating the importance of information on quality and in knowing how to use it. In addressing these two issues, the project also aims to develop and identify processes and technical systems which will support and encourage the appropriate use of data.

This paper will provide an overview of the presentation and education initiatives which have arisen from this project. The paper will then explore the ways different methods of presentation, the systems which support them and education strategies interact with each other. In particular, the paper will comment on the importance in supporting education strategies with well developed systems and appropriate presentation methods.

KEY WORDS: Data quality; Quality framework; Communicating Quality

1. INTRODUCTION

National Statistical Offices (NSO's) produce statistics to assist the government and wider community in making informed decisions. Thus, the issue of quality is fundamental - in order to use the data appropriately, the limitations of the data must be clearly understood. Similarly, the people need to understand how these limitations affect the decision-making process. This need to understand the quality of the data is becoming even more important as people try to extract more and more information out of the available data, particularly as NSO's continue to work towards improving the accessibility of their statistics, particularly in the field of electronic dissemination. This prompts two key questions:

- How do we make information on the quality of the data useable and freely available to those that use the data?
- How is the information on the quality of the data going to help people when they come to using the data?

Data are generally referenced with a particular use in mind. Understanding the quality issues associated with the data is important to ensure that the data are fit for the purposes for which they are intended. For example, peculiarities regarding the scope of the data (eg. only covers employing businesses) or the accuracy of the data (eg. high sampling error) may impact on the usefulness of the data for a particular user. Furthermore, knowing the potential limitations of the data also assists them to formulate and apply

---

[1] The views expressed in this paper are those of the author and do not necessarily reflect those of the
  Australian Bureau of Statistics.

[2] Bill Allen, Australian Bureau of Statistics, PO Box 10, Belconnen, ACT, 2616, Australia ,
  bill.allen@abs.gov.au

appropriate risk management strategies. As such, it is a combination of the data and the information on the quality of the data which forms the basis for informed decision making.

The Qualifying Quality project was set up in late 1999 in the context of addressing the following three key desired outcomes:

1. People appreciate the importance of knowing about the quality of the data they are using.
2. Information on the quality of the data is readily available to people for all data available from the Australian Bureau of Statistics (ABS).
3. People are able to use information about data quality to determine how well the data they are using meets their needs. This will often be incorporated into risk management.

The project focuses on these key outcomes through two key directions; presentation and education. Presentation relates to providing people with the information on the quality of the data and helps answer the question "Are the data fit for purpose". Education assists those people in appreciating the importance of information on quality and in knowing how to use it.   However, both presentation and education rely on the support of strong underlying processes and technical systems. Through establishing standard processes, such as automatically including information of quality with adhoc information consultancies, the importance of quality is reinforced. Similarly, through clear and focused presentation of this information on quality, users of the data are both reminded of the importance of understanding the quality of the data and helped towards the effective use of such information. As part of this, it is important that technical systems support these presentation and dissemination methods such that it is easy for staff within the NSO to integrate this focus on quality in their day-to-day work.

Having established that presentation, process and formal education all have important roles to play, it is also important to be aware of the way they interact. In particular, it is important the different aspects correctly support each other. In particular, formal education without systems to support any suggested changes in process will naturally struggle. As such, it will be important to ensure that as new strategies are adopted, processes are developed to support the strategies before trying to fully integrate them both inside and outside the ABS.

In addressing this need to develop underlying process and systems to support our initiatives, the Qualifying Quality project has included the development of technical system prototypes to aid in the presentation and dissemination of information on quality, and a formal training course focusing on the education of users of the data. Using the Inputs - Transformations - Outcomes (ITO) model developed for ABS (Smyrk, 1999), these current initiatives in the Qualifying Project can be broadly summarised as in Figure 1 below. More information on the ITO model is provided in Section 4.

Figure 1. Qualifying Quality Project

2
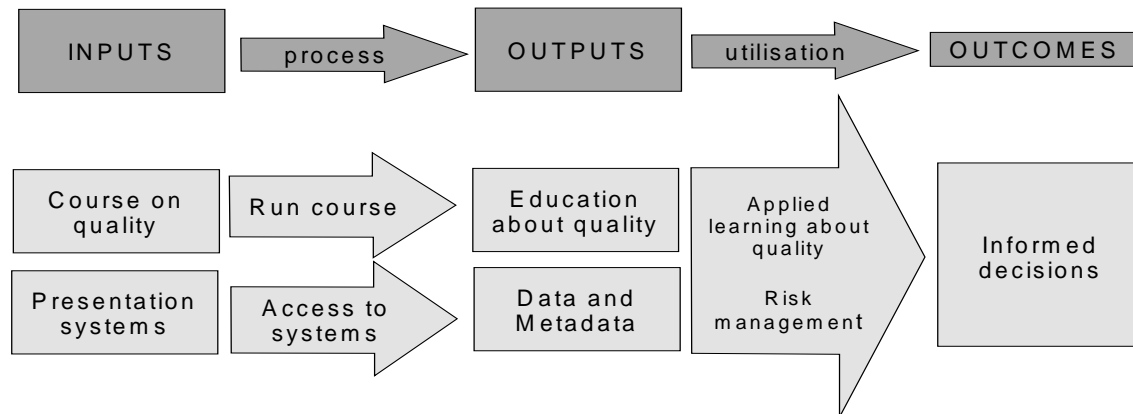
Finally, the project focuses on the general user of data who has been drawn to a dataset through 'headline' statistics and then decided to delve further into detailed tabulations to better understand the finer issues. These general users will usually be relatively uninformed about the quality of the data and are the key focus for these presentation and educational foci.

An overview of the current initiatives of the project has been recently provided in Lee and Allen (2001). This paper will focus specifically on the presentation and education issues in the context on communicating quality to these general users.

In Section 2, the paper will discuss the different roles data quality frameworks can take and how these roles relate to the general user. Section 3 will introduce the data quality framework being adopted in the Qualifying Quality project and provide an overview of the presentation systems prototypes from the Qualifying Quality project which draw on the data quality framework. Section 4 will then describe current work being done on developing a formal course based on applying the data quality framework. Finally, Section 5 will briefly cover other initiatives associated with the Qualifying Quality project.

2. THE ROLE OF DATA QUALITY FRAMEWORKS

There has already been significant comment on the change in focus for NSO's to the perspective of the user. Dobbs et al. (1998) and Brackstone (2000) note that this translates to a greater responsibility for NSO's to report on the quality of the data they disseminate. It is not realistic for a NSO to successfully meet all the needs of all potential users, so the ultimate assessment whether data are of high quality, that is whether data are fit for purpose, must lie with the user.

This has led to a number of agencies developing 'quality checklists' and adopting data quality frameworks. Typical examples of quality frameworks include those provided by Statistics Canada (Brackstone (2000)), the Korea National Statistical Office (Lee, D. and Shon, A. (2000)) and the International Monetary Fund (IMF) (Carson, C. S. (2001)). Lee, D. and Shon, A. (2000) also provide a comparison of the dimensions of quality for Canada, the Netherlands, Korea, EuroStat and the IMF, noting that while the dimensions of quality differ between countries, they are largely comparable with common themes of accuracy, relevance, timeliness and accessibility.

In considering the role of the frameworks, it is useful to first look at some typical uses of data in government: policy development and policy evaluation. In both cases, the data are a means to achieving some desired 'real world' outcome through appropriately focused policy. Thus data can be considered to act as windows to the real world - data are used to tell us what is really happening in the world. Extending the analogy further, quality declarations tell us how clear the window is or even whether the window is looking in the right direction!

The roles of these frameworks have been described as assisting in the production of reports on quality, either for internal quality improvement and quality assurance or for an external *quality declaration*, Elvers E. and Rosén B. (2000). In these instances, the advantage in using the frameworks has generally been described as assisting towards ensuring that all appropriate aspects of quality are covered in the reports. Thus, the frameworks act as a prompt for producing a more complete coverage of quality.

The other role which can tend to go unstated, although the benefits can be clearly realised, is that of education. That is, data quality frameworks are an invaluable tool in educating people (both inside the NSO and external users of data) about quality. More specifically, when used appropriately, data quality frameworks offer the following advantages:

- they provide a common ground for communication about quality;
- they provide the basis for a checklist to ensure a broader concept of quality is considered; and
- when linked with a systematic approach, data quality frameworks form a basis for improved integration of quality into the decision-making process.

In considering the role of data quality frameworks, it is useful to draw comparisons with the role of language in wording questionnaires. Clark, H. H. and Schober, M. F. (1992) make the observation that "... you can't understand what happens in survey interviews without understanding the role of intentions in language use". This is also true with data quality frameworks as the language and apparent intentions of the framework will influence the way in which the frameworks are used. More specifically, the data quality framework provides a common basis for discussing aspects relating to quality. This fits in with what Clark, H. H. and Schober, M. F. (1992) refer to as the Principle of utterance design:

"Speakers try to design each utterance so that their addressees can figure out what they mean by considering the utterance against their common ground".

This common ground can also be used to highlight the role of all aspects of quality. Quality in surveys is often referenced through the concept of total survey error (Lyberg L., Japec L. and Biemer P. (1998)), which expands upon the earlier measures which focused just on sampling error. The data quality framework both formalises the concept that all aspects of quality need to be considered, and provides an overview of the different aspects of quality. This overview can then be used as a basis for producing a 'quality checklist' which can be used to populate our quality declaration.

From an educational viewpoint, it must also be remembered that this common ground needs to cover both staff within the NSO and the general users of data. While it is big advantage for everyone within the NSO to speak the same language in evaluating quality, the benefits are even greater when the general user also speaks the same language. Thus, the general user is then better placed to query the data which can lead to improvements for both the user and the NSO - the user gains improved understanding of the data, while the NSO is better placed to identify areas for improvement, including both the data and the delivery of metadata.

Finally, the data quality framework can be integrated into a systematic process to provide people with a way to use information on quality in their day-to-day decision-making. In trying to educate people about the importance of quality, a key issue has to be that information on quality impacts on the types of decisions they make. Similarly, people need to understand that getting high quality data is not about purchasing high quality products which will automatically meet their needs, but rather depends on appropriately matching their needs with available data sources. The data quality framework plays a key role in addressing these educational needs and will be discussed further when a formal course being developed is discussed in a later section.

Before moving on, it is important to stress the importance of choosing a framework which is itself 'fit for purpose'. Clark, H. H. and Schober, M. F. (1992) also make the following key points with respects to survey questions but they equally apply to the choice of a data quality framework:

- *Respondents answer vaguely worded questions in idiosyncratic ways; and*
- *Respondents fail to see when the surveyer is using words differently from the way that they use them.*
- *The response alternatives to a question help determine the domain of inquiry in which it is to be answered.*

The first two points stress the importance of having a well-defined framework, touching on the concept of interpretability. Accordingly, the clarity and implied meanings within the specific words in the framework also need to be carefully selected - the more the words rely on cross-referencing to broader definitions, the less effective they become in assisting the general user. Similarly, the third point highlights the advantages in using a data quality framework. The framework helps to assist people though providing a set of issues to consider when making a decision, ensuring that all of the appropriate aspects of quality are covered.

3. DEVELOPMENTAL PROTOTYPES USING DATA QUALITY FRAMEWORKS

An important aspect of the Qualifying Quality project is the development of a number of prototypes which focus on ensuring that the presentation aspects of the project are adequately supported by underlying systems. These prototypes represent work in progress and are specifically to assess the methods being suggested, before any properly quality assured production systems are built. In addition, a formal course is being developed to support the education aspects of the project.

The development of two of these prototypes has drawn heavily of the roles of the data quality framework described above, as has the development of the formal course. In doing so, the Qualifying Quality project has adopted the framework proposed by Statistics Canada (Brackstone, 2000), covering the aspects of Relevance, Accuracy, Timeliness, Accessibility, Interpretability and Coherence.

These six aspects were considered appropriate both in their coverage of quality and in the degree to which the concepts could be readily understood by the general user. While some minor clarification is needed on what is meant by terms such as relevance and interpretability, the terms are expressed in relatively plain English and, as such, provide a good basis for building a common ground for general users to talk about data quality.

The two prototypes which reference the data quality framework have been broadly described as:

- Quality Issue Summaries; and (indirectly)
- Improved methods for the presentation of data with metadata.

The concept behind the Quality Issue Summary is to provide a short *summary* for each survey of the *issues* relating to *quality*. The intended audience is the general user who will use the Summary to quickly assess whether the data are likely to be fit for purpose. Thus, the Summary needs to be both short enough so that people will not be overly adverse to reading it and concise enough that the sufficient information is provided to assist the user to assess the fitness for purpose of the data.

The Quality Issue Summary can be thought of as describing the data from the viewpoint of the end user. This is distinct from what might be called a Quality Assessment, which describes the data from the viewpoint of the underlying processes which led to its creation. Rather, the Quality Issue Summary assists the end user to assess fitness for purpose, while the Quality Assessment plays a useful role for the NSO as part of its internal quality assurance and quality improvement processes.

Figure 2. Quality Assessments and Quality Issue Summaries
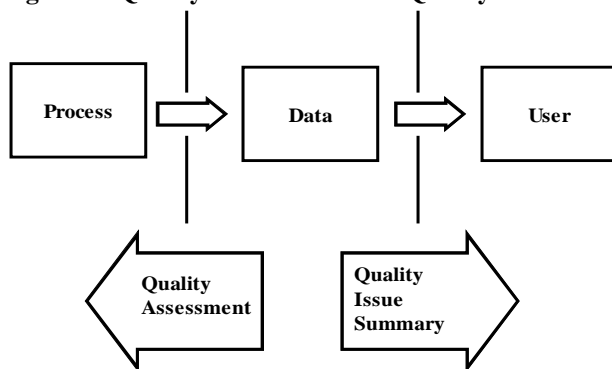
The data quality framework plays a key role for the Summary as it subdivides the Summary into six sections, with each section corresponding to a different aspect of quality. In doing so, it breaks up the summary into smaller, better defined issues which assists both the final user and the statistical organisation

populating the summaries. Similarly, it encourages people to consider a wider aspect of issues when thinking about the quality of data.

The second prototype indirectly references the data quality framework, as it accesses Quality Issue Summaries. This prototype focuses on using the advantages of an electronic medium to better integrate data and metadata within a single table. For example, dynamic links on row and column headings could readily provide metadata on the headings (e.g. classification information). Similarly, a dynamic link would provide easy access to the Quality Issue Summaries.

4. A FORMAL COURSE

The formal course also relies heavily on the use of a data quality framework. The course has been tentatively named "Using Quality to Make Better Decisions", although an improved name is being sought. It is currently projected to be one day long and be offered to external users, although people from the national statistical office would also benefit from attending.

The key focus of the course is to provide users with a systematic process for integrating information on quality into their decision-making process. Firstly, the course introduces the nature of quality as 'fit for purpose' and describes how a data quality framework can assist in defining quality. The course starts with the aim of making an informed decision based on data. The decision required at the end dictates the data need. The data need then focuses attention on certain data sources. These data sources are then used to feed back into the final decision.

Figure 3. Initial Learning Map
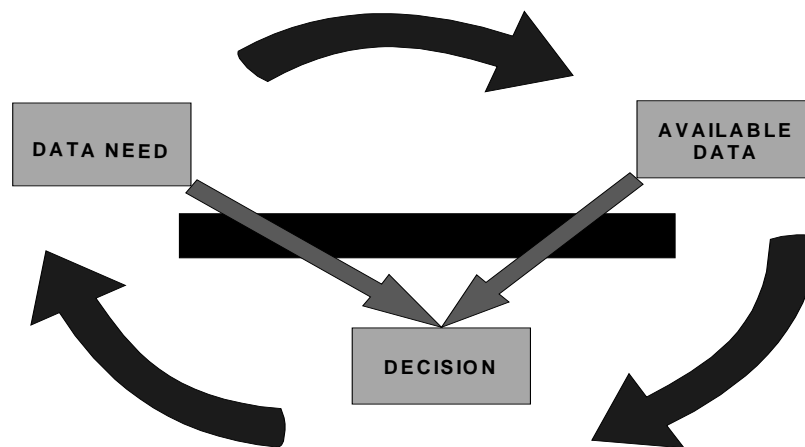
Next, the course touches on the issues of defining outputs, related to desired outcomes, ensuring that the purpose is well understood. This is achieved through referencing the Inputs - Transformations - Outcomes Model (Smyrk, 1999). The Inputs - Transformation - Outcomes (ITO) model defines a project structure in terms of its objectives (outcomes), its deliverables (outputs) and how the project inputs are transformed, via outputs, into outcomes. Outcomes are the ultimate objective, whereas the outputs are the physical deliverables that help achieve the outcomes. For example, the outcome for the ABS might be to assist in informed decision-making, whereas the outputs might include deliverable such as statistics in publications and on the ABS Website, and consultancy services. The inputs would then include user consultation, data collection and ABS staff and processes. For the course participants, the outcomes would often correspond to 'real-world impacts', with the outputs corresponding to policy implementation and evaluation . Inputs would correspond to policy development.

Participants are then encouraged to use the data quality framework developed by Statistics Canada to clearly define their data need - 'fit for purpose' implies a well-defined data need. People are guided through

defining their data need using the six aspects of quality. For example, participants are encouraged to list their relevant concepts, accuracy requirements, time constraints, and so on.

Having established their data need, the course recommends using the same process (and data quality framework) to assess the available data sources, leading to a natural comparison between the data need and data sources for each of the six aspects of quality. The Quality Issue Summaries described above are then introduced as a quality reference for ABS data. This session also illustrates that datasets are not suitable for all possible needs, reinforcing the concept of 'fit for purpose'. Similarly, the session introduces the importance of sourcing multiple datasets where possible.

The next session deals with comparing the data need and the available data sources. Firstly, simple examples are used to pick one dataset over another on the basis of a specific need. Then, the session looks at more complex methods of applying the results of our assessment of quality. More specifically, the focus is on integrating the information from the quality assessment into the decision-making process. The session introduces concepts such as sensitivity analysis, contingency planning, accessing multiple datasets and deciding on the need to create a data source. Sensitivity analysis is used to question participants how their decisions might change if the data were different. Contingency planning asks participants to consider what actions might be required if they find out their previous decisions have been made on the basis of misleading information. Accessing multiple datasets is used to introduce two issues: using different data sources to answer different parts of the question; and using multiple data sources to try and assess whether there is a consistent picture. Finally, deciding on collecting more information reinforces the fact that existing data sources may not meet your purposes sufficiently to make an informed decision.

Figure 4. Final Learning Map
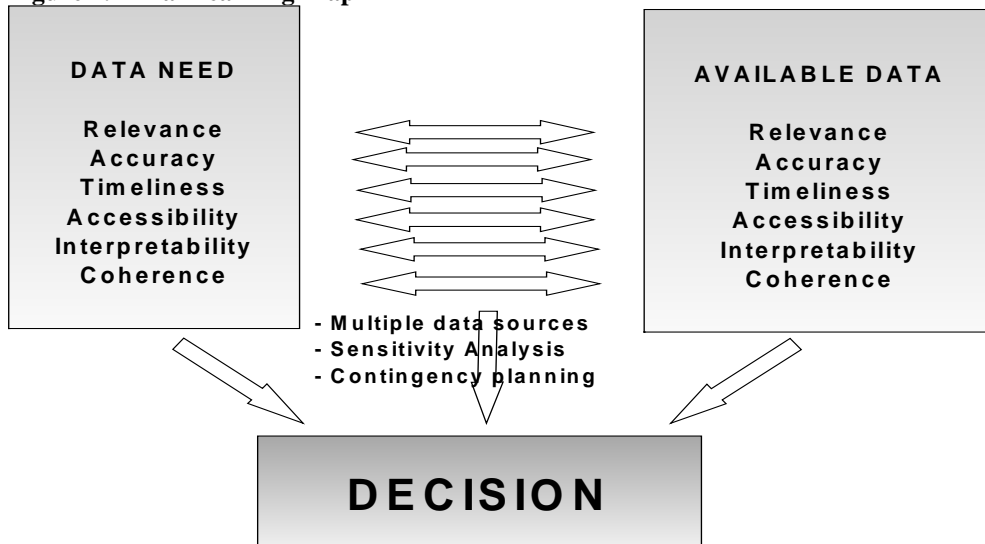
Finally, the course recaps on the issues presented and discusses how participants can apply these principles in their day-to-day work. The session also talks about using the principles more generally for evaluation and for the design and evaluation of performance indicators.

5. ADDITIONAL INITIATIVES

Additional initiatives are also being proposed or trialed, both within and outside the scope of the Qualifying Quality project.

As part of the project, two additional prototypes are being developed and a second course is being proposed. The first prototype focuses on the development of a system which provides improved access and presentation to stored (quantitative) quality measures. This prototype is initially targeted for use within the

ABS and provides for both tabular and chart presentations. External access to this system in the future will also be considered.

The second prototype focuses on providing more flexibility regarding presentation of data accuracy measures. While the prototype is initially looking at providing customised annotations to survey estimates according to estimated relative standard errors, the scope of the data accuracy initiative extends further. Over time, it is hoped that technical presentation systems will fully provide for standard errors and confidence intervals which should assist in their informed use within the user community. Further, there are some plans to eventually extend these presentation systems to include basic hypothesis testing.

Another formal course is also being considered, although it is not clear whether the course would be offered as an addition to the current course being developed, or as a stand-alone course. The course would focus on the same concept of applying a systematic approach using the data quality framework, although the focus would be on the development and assessment of performance indicators. Performance indicators can be considered as simply another data source and, as such, can be fully considered within the course being currently developed, as outlined in Section 4. However, it is envisaged that this additional course would both provide for a more specific focus for the application of the data quality framework and educate people who may not have seen the data-focused course as sufficiently appropriate to attend. As such, the course would lead to an increased appreciation and understanding of the role of data quality in the process of making informed decisions.

In addition, the Statistical Consultancy Unit (SCU) within the ABS has been experimenting with the data quality framework outside the scope of the Qualifying Quality Project. The SCU provides statistical consultancy services, primarily to official bodies on a cost-recovery basis. One of the key services offered by the SCU is the methodological review. This involves reviewing an existing or proposed survey, audit, analysis or other statistical process, considering the overall level of statistical rigour and noting the appropriateness of the methodology. This includes identifying underlying assumptions being made and detailing the strengths and weaknesses of the approaches being suggested or implemented. Thus, the methodological review can be considered to be an assessment of 'fit for purpose' from a methodological viewpoint. This is also similar to the advice provided for tender evaluations.

More recently, the Statistical Consultancy has applied the framework while conducting methodological reviews, including the assessment of performance indicators, data sources and the presentation of statistical data. The use of the framework has been generally well received by clients, particularly the Australian National Audit Office (ANAO), who have appreciated the direction and coverage provided by a formalised data quality framework. As a result, a recent audit report by the ANAO (2001) recommended the use of the data quality framework as 'better practice' in specifying performance measures.

REFERENCES

Australian National Audit Office (2001), "ATO Performance Reporting under the Outcomes and Outputs Framework, Australian Taxation Office", Audit Report No.46 2000-01, pp63-64.

Brackstone, G. (1999), "Managing Data Quality in a Statistical Agency", *Survey Methodology*, Vol.25, No. 2, pp. 139-149

Carson, C. S. (2001), "Towards a Framework for Assessing Data Quality", *IMF Working Paper*, WP/01/25

Clark, H. H. and Schober, M. F. (1992), "Asking Questions and Influencing Answers", in J. M. Tanur (ed.) *Questions about Questions, Inquiries into Cognitive Bases of Surveys*, New York: Russell Sage Foundation, pp. 15-48.

Dobbs, J., Gibbins, C., Martin, J., Davies, P. and Dodd, T. (1998), "Reporting on Data Quality and Process Quality", *Proceedings of the Survey Research Methods Section, American Statistical Society Association*, pp. 32-40

Elvers, E. and Rosén, B. (2000), "Quality Concept for Official Statistics", *Encyclopaedia of Statistical Sciences*, Update Volume 3, pp. 621-629.

Lee, D. and Shon, A. (2000), "Korea's Experiences in Statistical Quality Assessment", *Statistical Quality Seminar 2000, International Monetary Fund and Korean National Statistical Office*

Lee, G. and Allen, B. (2001), "Educated Use of Information about Data Quality", *53rd Session of International Statistical Institute*

Lyberg, L., Japec, L., and Biemer, P. (1998), "Quality Improvements in Surveys - A Process Perspective", *Proceedings of the Survey Research Methods Section, American Statistical Society Association*, pp. 23-31

Smyrk, J. (1999), "Project 'Solutions': Who is accountable?", *1999 National Conference of the Australian Institute of Project Management*