
Measuring to Improve Quality and Productivity in a Processing Environment

Jeri M. Mulrow, Internal Revenue Service and
Fritz J. Scheuren, George Washington University

The Statistics of Income (SOI) of the Internal Revenue Service (IRS) is charged with collecting data from tax returns for use by government and private researchers in the analysis of tax policy and the economy. Currently, SOI transcribes and edits, on an annual basis, over 500,000 tax returns in five processing centers across the nation. This paper attempts to link the challenges and changes in SOI processes over the years to Deming's Fourteen Points of quality improvement.

The first section of the paper provides a limited background on the Statistics of Income program. The second section touches on SOI statistical processes. In the third section, each of W. Edwards Deming's Fourteen Points (Deming, 1982) is discussed with regard to the progress that SOI *has* or *has not* made over the years -- on each process and each of Deming's points. A method to display these results in a simple yet overall way is presented. The next section describes a scoring method using a couple of examples from SOI. The last section ends with a few concluding remarks.

■ Background

Statistical operations at IRS began over 80 years ago, with the passage of the Sixteenth Amendment to the United States Constitution in 1913. Throughout its long history the Statistics of Income organization (and its predecessors) has had the key role of converting administrative data from tax returns into statistical information (e.g., Norwood, 1995). SOI's long tradition has engendered real pride in the work done and a strong feeling of knowing "how to do it well." This is both good and bad.

Pride in Workmanship is one of Deming's fourteen points, and SOI certainly has that. However, pride can sometimes be a hinderance to accepting new ways of doing things (Adopt the New Philoso-

phy), or learning new techniques (Institute Leadership, Education and Self-Improvement) or in striving to continually improve (Constancy of Purpose, Continuous Improvement) -- since we do "it" so well already.

The Statistics of Income programs are very product-driven. There are tight time frames to collect, clean, and distribute the data. There are specific customers who put requirements on the data. There are specific processes that must be completed in a fixed order. Work is treated much the same as if it were being done on a production line.

It is hard to see ourselves as providing a service, instead of a product. But information is a service. Statistics of Income personnel can not only provide the required data, but also can provide a wealth of expert economic and statistical knowledge in the use of those data. Sadly, this knowledge has not often been utilized or even recognized throughout much of SOI's long history.

The three major customers of the SOI data are the Office of Tax Analysis (OTA) in the U.S. Department of the Treasury, the Bureau of Economic Analysis (BEA) in the U.S. Department of the Commerce, and the Joint Committee on Taxation (JCT) in the U.S. Congress.

Unrestricted access to SOI microdata is only available, by law, to OTA and JCT; BEA has limited access to corporate data and all other users are confined mainly to aggregate statistics. With few exceptions, public-use files are generally not available, due to disclosure and confidentiality concerns (Strudler, Oh, and Scheuren, 1987).

Quality improvement has always been a concern of SOI. Over 30 years ago, for example, W. Edwards Deming was asked to a "Review of the

Sampling Procedures Used by the Internal Revenue Service to Produce Statistics of Income from Individual Tax Returns, with Special Emphasis on Achievement of Quality" (Deming, 1963). In his report, Deming offered many suggestions for improving the accuracy of the figures and results produced by SOI. However, he was not asked, and did not provide, instruction on how SOI (or IRS) could undertake the kind of transformation that the Japanese were making at that time to their processes. Nonetheless, Deming did discuss quality control, variability, the need for "guidance" of workers, continual feedback and balancing of resources -- many of the underlying topics in his 14 Points.

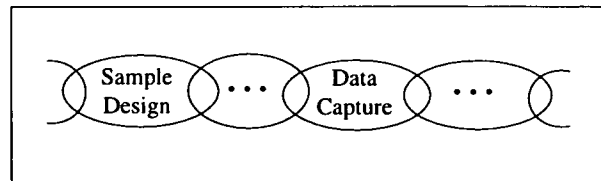
While much good work was done, still for over 20 years neither IRS generally, nor SOI specifically, moved much in the direction of "Adopting a New Philosophy." Not until the disastrous filing season of 1985, when the IRS actually lost whole shipments of tax returns and incurred sizable interest payments on late refunds to taxpayers, did the IRS make any steps in this direction. The 1985 failure put IRS under intense internal pressure to change. There was also great outside pressure from both Congress and even ordinary taxpayers to explain how such a thing could occur and to take steps towards preventing a recurrence.

In this crisis, after a discussion of Deming's approach, the IRS turned towards the teachings of J.M. Juran (Juran, 1964) and plunged into the modern quality arena. Throughout the next few years, IRS expended a lot of time and energy using Juran's ideas to improve its systems -- even winning some of the top quality awards in the Nation. Both the Cincinnati and Ogden Service Centers won the Office of Management and Budget's (OMB) Quality Improvement Prototype Award and Ogden went on to win the OMB Presidential Award for Quality in 1992. In 1989, the Statistics of Income organization also applied for (but did not win) the OMB Quality Improvement Prototype Award.

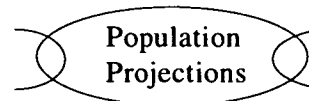
■ SOI Statistical Processes

Statistics of Income processes can be thought of as being linear and sequential, like a chain. Each process is dependent on the process before it, and right now they are performed in a sequential manner. Figure 1 shows this type of thinking.

Figure 1.--Statistical Processes as a Chain



Like a chain, SOI processing is only as strong as its weakest link. In this section, each of the major statistical processes are briefly described.



Most of the SOI studies are based on stratified probability samples of tax returns. There are nearly 60 separate studies, each with its own sample design, such as the Individual tax return study, the Corporation tax return study, and the Estate tax return study.

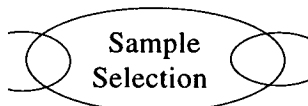
In most studies, the tax returns are sampled on a flow basis, as they are filed and worked within the IRS. This type of sampling allows for easier retrieval of the tax return for further statistical processing, but presents a bit of a problem for design. The final population counts are not known until the end of the sampling period. However, "fairly decent" estimates of the population totals can be made based on historic and other information. Unfortunately, due to the typical year-and-a-half lag-time between design and selection of the sample, population projections have to be based on two or three year old data. This lag is a direct consequence of the restrictions placed on the process to select returns from the IRS Master Files.



Sample Design

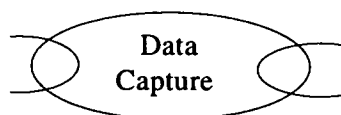
The current sample designs basically have been optimized to provide yearly cross-sectional estimates of population totals. As noted earlier, roughly 500,000 tax returns are sampled annually for statistical purposes from a population of over 200 million tax returns filed each year with the IRS. Many of the SOI studies also have embedded longitudinal data as a result of the sampling mechanism used by SOI (e.g., Hinkins, Mulrow, and Collins, 1990; Hinkins and Hughes, 1995).

The sampling technique uses a taxpayer's unique identification number (TIN) as the seed for a pseudo-random number (Harte, 1986; Sunter, 1986). Since the TIN typically remains the same from year to year, a taxpayer selected for the sample in one year has a high probability of selection in the sample in the next year, barring any large changes in economic status of the taxpayer from one year to the next. Modifications to the sampling strata will have a slight effect on the selection of the taxpayer.



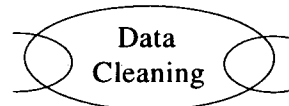
Sample Selection

Tax returns are filed year-round in ten different IRS Service Centers scattered throughout the country. The Service Centers feed tax administration information, on a weekly basis, to the IRS Master File systems, where the sample selection process for SOI studies occurs. Sample selection information is then fed back to the individual Service Centers where the tax returns are physically located. The sampled tax returns are manually retrieved from the files and set aside for statistical processing. In the case of electronically filed returns, a copy of the electronic data is transferred to SOI for statistical processing.



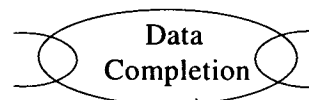
Data Capture

Statistical data capture is performed in one to five of the IRS Service Centers, depending on the SOI study. The Statistics of Income studies provide many more additional data items than are available administratively, including information from nearly all of the forms and schedules filed with the tax returns. SOI data are collected in a relational database system using on-line transaction processing. All data capture operations are now completed in one pass, including key-entering and error correction. Reduced handling costs, greater accountability, and a sense of ownership have resulted.



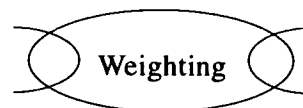
Data Cleaning

Due to substantial penalties for misreporting, the detailed income and expenditure data on tax returns are generally regarded as more reliable than similar survey data. Even so, SOI goes to great lengths to protect against nonsampling errors, such as those due to the taxpayer or data processing errors. Extensive on-line tests for consistency and reliability, based on tax law and the improbability of certain data combinations, are employed. Further, subsamples of the work are independently processed and compared as a check.



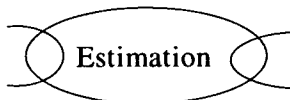
Data Completion

Missing data, either at the item level or at the unit level, occur very infrequently. Most of the missing data problems are solved using a combination of weighting and/or imputation techniques. For example, an estimate can often be based on other data from the tax return or schedule, or prior-year data for the same or a similar taxpayer, or same year data from "similar" returns (e.g., Überall, 1995). Multiple imputation techniques have proven highly successful (Rubin, 1987).

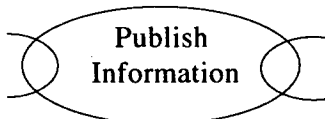


Weighting

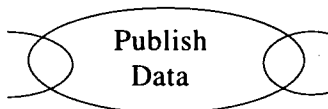
For most of the SOI studies, the weighting procedure is fairly simple. The weights are calculated by dividing the final population counts for each stratum by the corresponding final sample count for that stratum. In some of the studies, the estimates are improved by using post-stratification and raking techniques (e.g., Mulrow, Oh, and Collins, 1991).



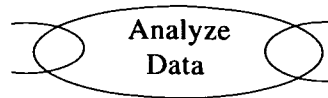
The final data estimates are straightforward. Most released data are aggregate population estimates for specific subdomains of interest. Recently, more of the studies have been providing advanced data estimates based on partial samples. These estimates require additional thought and statistical techniques, such as augmenting with survey data or using propensity scoring (Schirm and Czajka, 1993).



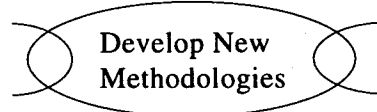
Aggregate statistics are published annually for most of the SOI studies. Some studies publish entire volumes of data, while other studies release data in the *Statistics of Income Bulletin*. Aggregate level data are available in paper, magnetic tape, diskette and electronic format via the SOI electronic bulletin board. Some will also be available soon on CD-ROM.



There are important disclosure and confidentiality concerns when publishing statistical information. For example, more research is needed in the areas of masking data and simulating microdata. Concerns from taxpayers at having such types of data available must be addressed (e.g., Duncan and Pearson, 1991). A limited effort by SOI has been made in this area in a joint study with the Bureau of Labor Statistics' Behavioral Science Research Laboratory.



Most of the analysis of the SOI data is done by outside users. A limited amount of data analysis is required to update or modify sample designs yearly and to produce advance data estimates. Traditionally, though, Statistics of Income analysts have looked at SOI data mainly as part of product delivery, not for process improvement. More SOI analytic emphasis generally -- for example, on good graphical techniques -- seems needed for large SOI data systems.



New methodologies are developed on an as-needed basis, not in a systematic fashion. They are often product-driven or customer requested. The old "If it ain't broke, don't fix it" attitude is prevalent. A lot more systematic work in this area is needed.

As the potential of new technology continues to grow and customer demands increase, the Statistics of Income programs will be forced to rethink their traditional notion of sequential processing. Parallel processing will have to become the wave of the future. Real time data estimates should become a realistic goal. The bindings that can hold these parallel processes together are Deming's 14 Points. This thought can be illustrated as in Figure 2.

Figure 2.--Parallel Processing

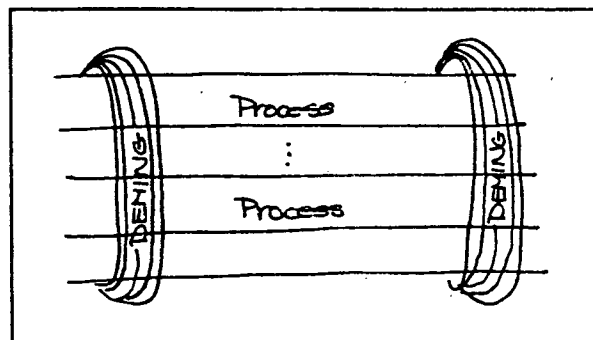


Figure 3.--Deming's 14 Points in Scherkenbach's Order

Deming Point	Essence	Icon
1. Constancy of Purpose	Strive to reduce the spread around the course (Set the course now for the future tomorrow)	
2. Adopt New Philosophy	Change the way we think about our business (No more business as usual)	
3. Design for Quality	Use processes for prevention of defects (“Cease dependence on mass inspection”)	
4. Continuous Improvement	Continue to reduce the variability (The customer's needs are a moving target)	
5. Pride in Work	Enjoy the work (Happy, satisfied workers perform the best)	
6. Drive out Fear	Require an atmosphere of mutual respect (Everyone must respect everyone else for their differing skills)	
7. Foster Teamwork	Work together to meet the customer's needs (Cross functional teams are a must)	
8. Eliminate Slogans and 9. Work Standards	Separate hype from purpose -- Focus on process not outcome	
10. Institute Leadership	“Follow me!” (Management must be constantly active in the process)	
11. Training and 12. Education	Fully understand the job (Need to understand the why and the how)	
13. Minimize Total Costs	Establish long-term relationships with suppliers	
14. Accomplish the Transformation	Transform all areas of our business (It is everyone's job)	

Deming has laid down a foundation for transformation that will be required for survival, including for SOI. The next section of this paper touches on each of his 14 Points.

■ Deming's 14 Points

Deming has presented his 14 Points over the years in many forms (e.g., Deming, 1982). Scherkenbach (1986) rearranged the order of Deming's points in *The Deming Route to Quality and Productivity: Road Maps and Roadblocks*. This paper will follow Scherkenbach's ordering (see Figure 3 on page 129). In that table, each of Deming's points is listed along with an icon or picture (some from Scherkenbach) that is intended to capture their essence.

For each of the 12 SOI processing steps, an assessment can be made of the progress made on Deming's 14 Points. Thinking in two dimensions leads naturally to a matrix presentation, where the rows of the matrix represent the processes and the columns represent the quality categories. Each cell of the matrix can be separately evaluated based on some type of scoring system (as below).

Figure 4.--IRS Statistics of Income Scoring

0	Extremely little or no action
1	Thought about it / no action
2	Action just beginning
3	Starting on the road
4	Firmly on the road
5	Becoming systematic
6	Systematic, beginning refinements
	•
	•
	•
10	Sound, systematic prevention, through continuous improvement

The magnitudes of the scores can then be plotted above the grid of the matrix, to create a three-dimensional response surface. The peaks on the

response surface represent areas where improvements are being or have been made. On the other hand, the valleys represent areas where little or nothing has been done. Such a surface provides, at a glance, a pictorial representation of how well the SOI organization is performing on its major processes. It also directly integrates processes with quality efforts.

In general, the surface above the matrix need not be smooth. It could and does have sharp cliffs and deep valleys. Potentially, as the surface changes shape, we obtain a history over time of how SOI operations are improving together (or separately).

Simply coloring in the squares of the matrix with different colors to denote different scores presents another way to easily use the matrix idea. The colors can be thought of as in a temperature gradient map. Certain colors, such as red, can represent hot spots of activity, while other colors, such as blue, can denote cooler areas of less or no activity.

■ Obtaining the Scores

A suggested scoring system that IRS Statistics of Income is using in this matrix evaluation method is a simplified version of the Malcolm Baldrige National Quality Award (MBNQA) scoring. Instead of scores ranging from 0 to 100 percent, though, the scores have been simplified to vary from 0 to 10. As SOI becomes more sophisticated at its quality efforts, the scale can be expanded back to the original scale of 0 to 100. For now, however, a rough order of magnitude measure should suffice.

A method is needed to help determine a score for each cell in the 12 by 14 array of SOI processing steps by Deming's 14 Points. We decided to use the format provided by Breheny (1995).

Breheny suggests laying out the vision, including the reason for change and what will change; the plan and tools to get there; and the

journey taken so far. For this paper, we have gone further and laid out some next steps. Two examples of specific cells might be worth sharing (see Figure 5 on page 132).

	Design for Quality
Data Cleaning	

The IRS Statistics of Income **vision** for this cell is to accurately capture tax law data for revenue estimation and tax law proposals. The **reason to change** is that a new processing environment was created for the entire data capture/cleaning process. The **result of the change** is expected to be less work at the end to complete the data files. The **plan** is to incorporate on-line consistency checks during the data capture and to develop an on-line quality review system. The **tools** to be provided to implement the plan are on-line help with problem resolution, internal benchmarking across programs, and developing a double edit system to catch discrepancies during review. The **journey so far** has allowed SOI to cut the time to produce a clean data record down from days to hours, even minutes in some cases and programs. On-line consistency tests have been built in and are continuing to be modified, as needed. On-line help is available for easy access and is continuing to be updated. The sample double edit system has been developed but not fully implemented nor fully utilized. The **next steps** are to continue to update and implement the improvements; and to look for new ideas outside IRS, such as at the Data Editing Workshop and Exhibition to be held in March 1996 in Washington, DC. The overall score for this cell was a **three**.

	Minimize Total Costs
Data Completion	

The **vision**, in this second example, is to provide accurate and timely data in the most efficient manner possible. The **reason for change** is to reduce the variability in the data due to many suppliers. The number of suppliers will be reduced, which should result in better data quality. The **plan** is to consolidate the number of IRS Service Centers that supply data. So far **on the journey**, SOI has consolidated from 10 suppliers to 5 or fewer, depending on the particular program. The **next steps** are to continue to rethink the data file production and to streamline the work, while piloting the delivery of real time data. The overall score for this cell was a **two**.

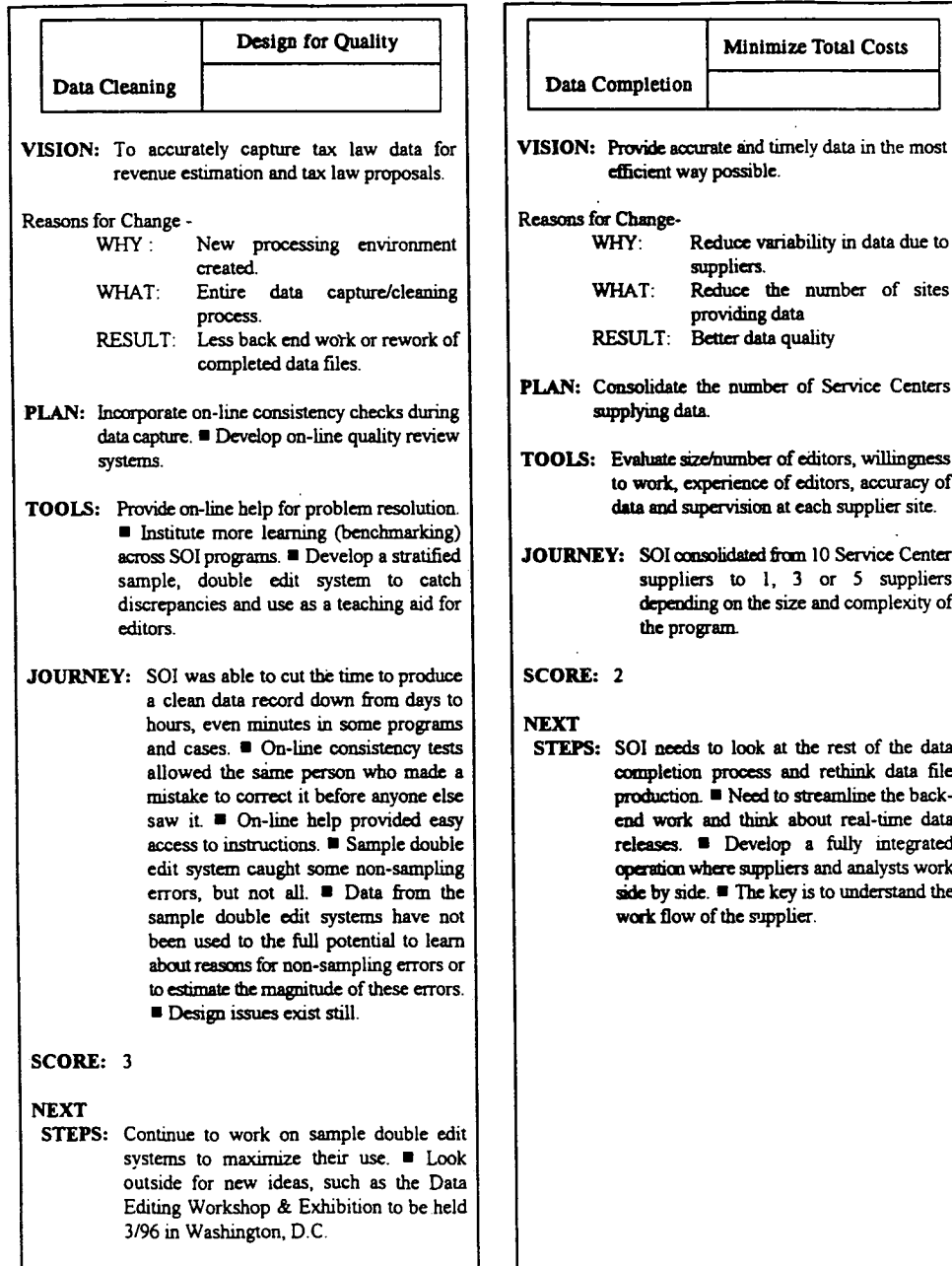
Anyway, systematically using this evaluation method, a complete matrix of scores was developed; it is given in Figure 6 (see page 133). In that figure, because of the limitations of these *Proceedings* (to two dimensions; black and white), we have chosen a third way of helping visualize the surface -- by sizing the scores to their magnitude.

The marginals of Figure 6 can also be evaluated. These marginals provide one-dimensional statistics on either the processes or the quality efforts. Figure 7 presents both marginals for the modified MBNQA scores -- in the form of bar graphs. Education and Training seems to be an area where SOI has made some progress. Likewise, the design, capture and cleaning processes score the highest. There is still a lot of room for improvement on both marginals.

■ **Concluding Comments**

In this paper a matrix evaluation method was introduced. It seems to have merit over other, say, linear evaluation methods. Clearly, it provides a natural way to integrate processes with quality efforts. Also, it gives a fairly fine grained assessment. Yet it is possible to get an overall assessment of the progress **at a glance**. Areas needing improvement can be targeted easily, while mak-

Figure 5.--Example of Breheny's Format



ing visible areas where improvements have already been made. The marginals can be used to shows which quality areas and which processes need more work. Finally, it is simple and straightforward to apply.

At the beginning of this paper, we talked about

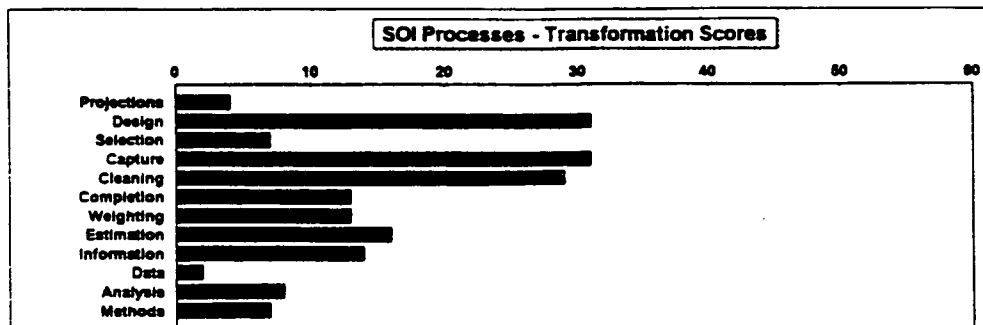
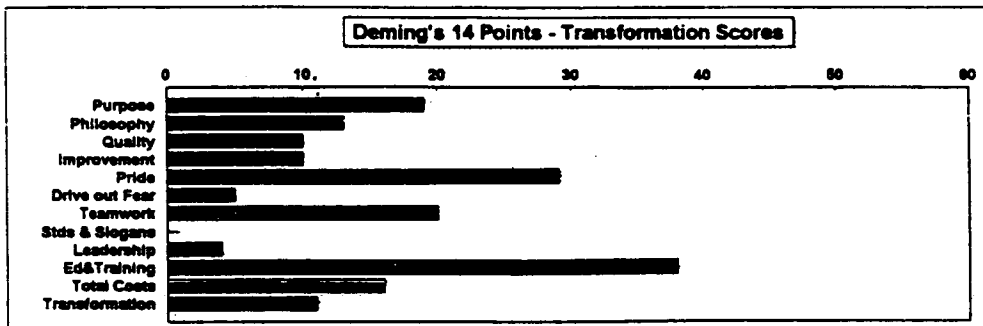
Deming's role with the SOI program -- some 30+ years ago. Clearly, his ideas are still being used in SOI today. A look at Figure 6, though, shows how much remains to be done. Perhaps the length of the journey yet to be traveled will discourage some. There have been many false steps; and, for those who keep going, there will be many more.

Figure 6.--Matrix Evaluation for IRS Statistics of Income, August 1995

Statistical Processes

Deming's 14 Points (SHERKELBACK'S ORDER)	Population Projections	Sample Design	Sample Selection	Data Capture	Data Cleaning	Data Completion	Weighting	Estimation	Publish Information	Publish Data	Data Analysis	New Methods
Purpose	.	4	1	4	4	1	1	2	1	.	.	1
Philosophy	.	4	.	3	3	.	1	1	1	.	.	.
Quality	.	1	1	3	3	1	.	.	1	.	.	.
Improvement	.	2	.	2	2	1	.	1	1	.	1	.
Pride	1	3	1	5	5	4	3	2	3	.	2	.
Drive out Fear	.	2	.	1	1	.	.	1
Teamwork	.	4	1	3	3	1	3	2	1	.	1	1
Stds & Slogans
Leadership	.	1	.	1	1	.	.	1
Ed&Training	2	5	2	3	3	2	4	4	3	2	3	5
Total Costs	1	3	1	4	2	2	.	1	2	.	.	.
Transformation	.	2	.	2	2	1	1	1	1	.	1	.

Figure 7.--Marginal Distributions from Matrix Evaluation



■ Acknowledgments

The authors would like to thank Lynne Stokes for her helpful discussant remarks. Mary Batchner too, needs thanks for organizing and chairing the session at which this paper was given. She is an individual that many will follow on the long journey ahead. We both are honored to know her and the other members of the SOI organization.

■ References

- Breheney, M. (1994). Finding the On-Ramp to Deming's Road, *Quality Progress*, 27, 10, 73-75.
- Deming, W.E. (1963). Review of the Sampling Procedures Used by the Internal Revenue Service to Produce Statistics of Income from Individual Tax Returns, with Special Emphasis on Achievement of Quality, reprinted in *Statistics of Income: Turning Administrative Systems Into Information Systems - 1994*, Internal Revenue Service (1995), 75-112.
- Deming, W.E. (1982). *Out of the Crisis*, Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, MA.
- Duncan, G. T. and Pearson, R. W. (1991). Enhancing Access to Microdata While Protecting Confidentiality: Prospects for the Future, *Statistical Science*, 6, 3, 219-239.
- Harte, J.M. (1986). Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS, *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Hinkins, S. and Hughes, E.S. (1995). Creation of Panel Data from Cross-Sectional Survey, *Proceedings of the Section on Survey Research Methods, American Statistical Association*. (Also in this volume.)
- Hinkins, S.; Mulrow, J.M.; and Collins, R. (1990). Design and Use of an Embedded Panel in the SOI Corporate Sample, *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Juran, J.M. (1964). *Managerial Breakthrough*, McGraw-Hill, New York.
- Mulrow, J.M.; Oh, H.L.; and Collins, R. (1991). An Evaluation of Bounded Raking Ratio Estimation in the Statistics of Income Corporate Programs, *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Norwood, J. L. (1995). *Organizing to Count*, The Urban Institute Press, Washington, DC.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.
- Scherkenbach, W.W. (1986). *The Deming Route to Quality and Productivity*, CeePress Books, Washington, DC.
- Scheuren, F. and Petska, T. (1993). Turning Administrative Systems Into Information Systems, *Journal of Official Statistics*, 9, 1, 109-119.
- Schirm A.L. and Czajka, J.L. (1993). Weighting a Panel of Individual Tax Returns for Cross-Sectional Estimation, *Statistics of Income: Turning Administrative Systems into Information Systems -- 1991-1992*, Internal Revenue Service, 145-154.
- Strudler, M.; Oh, H.L.; and Scheuren, F. (1987). Protection of Taxpayer Confidentiality with Respect to the Tax Model, *Statistics of Income and Related Administrative Research: 1986-1987*, Internal Revenue Service, Pub. 1299, 279-286.

Sunter, A.B. (1986). Implicit Longitudinal Sampling from Administrative Files: A Useful Technique, *Journal of Official Statistics*, 2, 2.

Uberall, B. (1995). Imputation of Balance Sheets for the 1992 SOI Corporate Program, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

■ Bibliography

Bernowski, K. (1995). What Makes American Teams Tick? *Quality Progress*, 28, 1, 39-44.

Bernowski, K. (1995). TQM: Flimsy Footing or Firm Foundation, *Quality Progress*, 28, 7, 27-30.

Boardman, T.J. (1994). The Statistician Who Changed the World: W. Edwards Deming, 1900-1993, *The American Statistician*, Aug. 1994, 49, 3, 179-187.

Corrigan, J.P. (1995). The Art of TQM, *Quality Progress*, 28, 7, 61-71.

DeToro, I. (1995). The 10 Pitfalls of Benchmarking, *Quality Progress*, Jan. 95, 29, 1, 61-63.

Dobbins, R.D. (1995). A Failure of Methods, Not Philosophy, *Quality Progress*, 28, 7, 31-34.

Fienberg, S. (1995). Overcoming the Real Issues of TQM Implementation, *Quality Progress*, 28, 7, 79-81.

Gitlow, H.S. (1994). A Comparison of Japanese Total Quality Control and Deming's Theory of Management, *The American Statistician*, Aug. 1994, 48, 3, 197-203.

Granquist, L. (1984). On the Role of Editing, *Statistisk tidskrift*, 2, 105 - 118.

Hare, L.B.; Hoerl, R.W.; Hromi, J.D.; and Snee, R.D. (1995). The Role of Statistical Thinking in Management, *Quality Progress*, 28, 2, 53-60.

Kivenko, K. (1994). Improve Performance by Driving Out Fear, *Quality Progress*, 27, 10, 77-80.

Landes, L. (1995). Leading the Duck at Mission Control, *Quality Progress*, 28, 7, 43-48.

Mann, N.R. (1994). In Memoriam: W. Edwards Deming 1900-1993, *Journal of the American Statistical Association*, 89, 426, 365-366.

Womack, J.P.; Jones, D.T.; and Roos, D. (1990). *The Machine that Changed the World*, Harper Perennial, New York, NY. ■