QUALITY ASSURANCE IN PRICE INDEX PROGRAMS

Eric D. Dmytrow, Bureau of Labor Statistics

Introduction

Managers in government and industry are concerned about quality. It seems that the media is full of Japanese success stories concerning product quality. The users of economic statistics want to know more about the quality of the statistics. What is quality? How do we achieve it? Why is it important?

The answers to these questions can be found in the approach to defect prevention and organizational excellence characterized as the management of quality. This approach emphasizes the managerial role in achieving quality and the usefulness of techniques like statistical quality control as tools.

I do not discuss the full implications of the management of quality or even attempt to address all of the questions above. I also will not discuss control charts, operating characteristics (oc) curves, or sampling plans. All of these are involved in achieving our results. I do want to suggest a conceptual model of survey quality and suggest a place that we can address the issues (dimensions) of quality in survey data.

Let me try to answer my last question with an example. Why is the quality of survey data important?

An economic statistic like the Consumer Price Index (CPI) or the Producer Price Index (PPI) has major influence on policy decisions and Federal as well as private expenditures. For example, the CPI is used to escalate many Federal entitlement programs. A one percent change in the CPI can generate approximately $2.5 billion in additional Federal outlays and at least that much additional outlay in the private sector.

What is Quality?

What is the most important aspect of producing economic statistics? Is it the economic conceptual design? The translation of design specifications into forms, procedures, and operations? The statistical methodology used to achieve the program design? Or is it the execution of the processes that we develop to satisfy the conceptual design and the statistical methodology? Well, I would answer that all of these are critical to the quality of the economic statistics survey. The most important area must be the conceptual design and the extent to which it satisfies the end uses of the data. The next important area is the statistical methodology (sample design) to achieve the end use requirements (the conceptual design). Finally, the execution of the processes (or the production activities, as I call them) are the least important of the three areas; however, poor execution of production or inadequate procedures can ruin a perfect conceptual design or statistical methodology.

I am going to discuss the execution of the production processes. Quality from this perspective is conformance to requirements. First I will elaborate more about the approach, how it is only one dimension of quality and how it fits into the overall management of the quality function. Then I will illustrate the results with four examples of some of the work done in the PPI.

The Production Activities

Consider the problem of producing "quality" economic statistics (price indexes) to be producing a "quality" information service--with fixed resources (people and dollars). Then our job as economists, as statisticians--as planners, and as managers--is to optimize that information service within the available fixed resources. The optimum resource allocation is important to achieve quality and the ultimate index quality depends on the three areas we have already discussed:

Conceptual Design
Statistical Methodology
Processing

The last area, the production activities, consumes most of our people resources. In fact, ninety percent of our people resources are consumed in production. What are these production processes? We have defined all the activities required to produce a monthly price index as "production." These production activities do not include design, or research and development activities. The six major production processes for the Producer Price Index Revision (PPIR) listed below:

Sampling Frame
Sample
Data Collection (initiation)
Monthly Repricing and Data Capture
Estimation
Publication

These six processes can be further expanded as in Figure 1 below.

Prior Work Examined
Process Control in Statistical Surveys

Can we not use some of the same science and rigor we use to measure economic phenomena to measure and control the processes that produce economic measures?

The answer is we can and we have. Statistical quality control has been practiced in large scale data entry processes for many years. Other statistical techniques have been used by survey organizations to measure and control processes involved in survey work. Two retired Census employees, George Minton and Herman Fasteau, worked with us when we were beginning to consolidate the quality control effort in the Producer Price Index Revision. George Minton (1970) and Herman Fasteau had pioneered some statistical quality control applications to the administrative

Figure 1

STAGE IN THE PRODUCTION OF THE PRODUCER PRICE INDEX

SAMPLING FRAME	SAMPLE	INITIATION	REPRICING	ESTIMATION	PUBLICATION
-Construct 4-digit SIC Listing	-Select industries to be indexed	-Interview sampled establishments	-Monthly updates of price and product specification data	-Calculate Indexes	-Monthly press release
-Update & refine Universe	-Prepare industry study	-Review collection forms	-Season and base in price data		-Monthly reports
	-Select and refine industry sample	-Transmit initiation packages to National office			-Annual reports
	-Prepare initiation materials	-Log in collection forms			-Articles
	-Transmit initiation package to Regional Offices	-Capture collected data in collection data base and produce review listings for DIPPI and SMD			
		-Release establishments to the Repricing System and the main data base			

environment. Minton and Fasteau went further than just establishing control--they measured the process yields and used the information to identify and eliminate the major sources of error. We have been doing that in the price index programs.

Quality Control Lessons from Manufacturing

Can we learn anything from the experience with quality control in manufacturing industries? The problems in manufacturing are obviously different than the ones we face in economic statistics surveys. However, there is much similarity in the issues of how to organize and manage a large group to achieve to goal. Large organizations have similar problems.

Many people have written books on the subject of quality control and the management of quality. Among them are Shewhart (1931), Deming (1960, 1982), Juran and Gryna (1980), Feigenbaum (1961), Crosby (1979) and others.

These books have three things in common:

1. There is a body of quality assurance principles for controlling and improving product/service quality from any process.
2. These principles are founded on empirical (and statistical) techniques.
3. The responsibility for applying these principles and techniques rests with top management.

Results from the Producer Price Index Organization Environment

The approach we have taken is to adopt the principles from quality assurance and try to create an environment that will offer the highest probability of success. This environment is characterized by certain elements. Some of these elements are listed below. The most important is the message from top management. Top management takes the leadership of the quality function and makes quality equally important to budget, schedules, and production. The other elements include the following:

Participation
Organization for improvement
Individuals are responsible for quality
Problems are recognized as management-controllable
Factual approach is used
Environment of blame is discouraged
Elimination of defect cause is critical

Participation means establishing an interdisciplinary approach to actions and involving the people doing the work.

Organization for improvement requires active, aggressive effort to make quality improvement happen, assigning specific projects to interdisciplinary groups.

Individuals are responsible for their own quality when processes are free of major management-controllable defects and individuals are provided with continuous information about how they are performing.

Most defects are management-controllable. The evidence from our studies supports this claim.

Decisions affecting quality must be made using factual information. Objectives measures and empirical data enhance management decisions.

Management's role is to discourage blame in the organization. If we subscribe to the premise.

The primary objective is to eliminate the source of the defect and to first concentrate on the major contributors of defects.

The Dimensions of Quality

How do we determine what quality is in a price index program? "Quality means fitness for use" says Juran (1980). Fitness for use reflects the notion that we need to look at the price index from the point of view of the users of the data. In order to achieve fitness for use we must address quality in the program through four major areas.

These four areas encompass all of the fitness for use criteria:

1. Design
2. Measurement
3. Conformance
4. Audit

Quality of design addresses how well the conceptual model represents the users' needs.

Quality measurement addresses how well the final index represents the design.

Quality of conformance addresses how well the organization executes the specifications and procedures.

Quality audit addresses how well the control processes are executed according to control specification.

If we reflect on our six major process steps (from Figure 1) in terms of these four areas of quality, we can construct a matrix (Figure 2). Appropriate quality efforts should be focused on each area of quality within each process step.

Figure 2 The Dalton Matrix

	Sampling Frame	Sample	Initiation	Repricing	Estimation	Publication
Design						
Measurement						
Conformance						
Audit						

The Dalton matrix illustrates the dimensions of quality and the systematic nature of the production process. This suggests an approach, at least to the issue of production quality. The approach we chose requires (1) studying the process yield in terms of conformance to requirements, (2) identifying the chief sources of error, (3) taking corrective action (if necessary), and (4) establishing quality control (measures). The projects follow reflect this approach.

The Disaggregation Study

The disaggregation study project that was done in the PPIR is a good example of the approach. Disaggregation is the last stage probability sampling technique used for selected items in the price programs. The first stage of the sample design in the PPIR is the selection of the particular companies and establishments whose cooperation will be requested. This stage is executed in Washington. The second stage (disaggregation) is the selection of a unique product and transaction that will be priced over time. This stage is executed by the BLS field collection staff in the establishments during the initial interview. Repricing is the monthly pricing of selected items after the initial interview.

Disaggregation consists of three basic steps:

1. Form a list of broad product classes (that includes all of the revenue sources).
2. Assign a measure of size to each class (preferably a revenue measure).
3. Select a product and transaction from these classes using probability proportional to size sampling techniques and repeat the process until a unique product and transaction are reached.

In the disaggregation study, we first measured conformance to data collection (disaggregation) procedures. Then we verified the procedures (specifications) against the PPIR design. We used the conformance measures to determine the process yield. The conceptual design (quality of design) did not change as a result of this work. However, we did satisfy ourselves that we could meet the design.

The basic conformance measure was the disaggregation score. The score ranged from zero to 100. Zero indicates no probability sampling was used. One-hundred is a case of perfect disaggregation, following the preferred procedures at every step. The score is calculated by summing the products of an assigned value for the prescribed procedure and the level of product detail arrived at by using the procedure. The assigned values reflect the relative importance of procedures to the desired outcome. The score measures how well the field representative was able to adhere to the preferred procedures. Perfect disaggregation, that is, using actual company records to scientifically select a unique product and transaction would result in a score of 100.

The scores formed a distribution skewed to the right. This suggests that for the most part scores above 60 were achievable. Sixty had been set as our minimally acceptable score. The frequency distribution (Figure 3) for one industry, is typical of the 11 industries studied.

There are several factors that can affect the interviewer's opportunity to conform to the procedures, but the data show that on average an acceptable level of conformance was a reasonable expectation for most industries. Conformance does not tell us how well the process is working in terms of the design. Are we achieving the desired distribution of products in the sample?

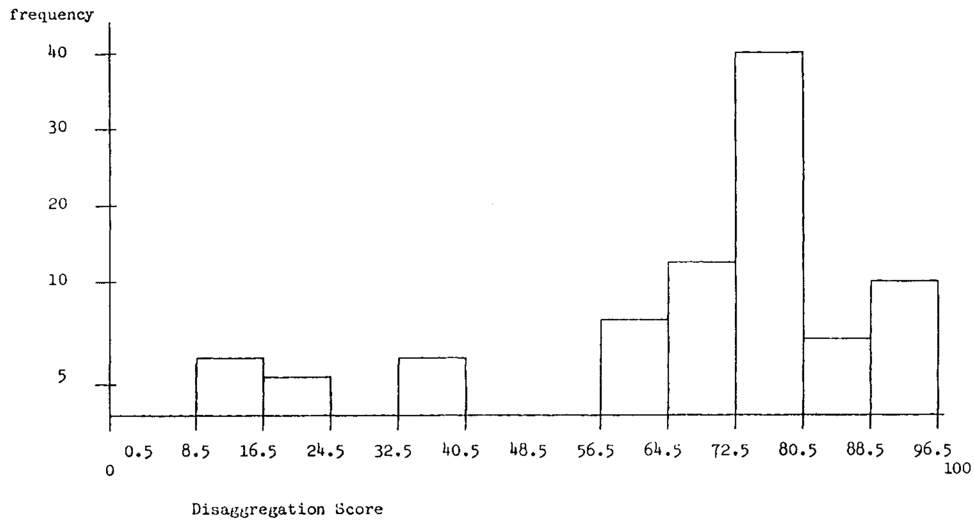
Gillespie (1981) compared the distribution of products in the sample to the distribution of products in the Census of Manufactures. The results indicated that where the disaggregation scores were the highest the sample product distributions most closely corresponded to the benchmark (Census) data.

We concluded from this work that, where the disaggregation scores were above 60 (on average), we achieved the desired sample.

Table 1 below depicts another way to analyze the data. The matrix shows the field representative by the industry. The symbols indicate whether the average scores were "acceptable" 60 or above or not (below 60). The patterns are revealing. Some field representatives had most (about two-thirds) of their cases below the acceptable average score. In fact only 17 percent of the field representatives accounted for over one-third of such cases.

At the same time, certain industries appear to be difficult for everyone. Industries C and G both exhibit that pattern. In fact 40 percent of the below standard cases are accounted for in these two industries (only 18

Figure 3



Disaggregation Scores by Field Representative and Industry

Table 1

Field Representative

Industry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Number of Price Quotes Collected	Average Disaggregation Score
A	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0	0	x	x	0	0	x	x	0	x				461	66
B	x		x							x								x	x	0	0		x		0		x			264	64
C	x	x	0				0	0	0	x			0		x	x	0				0	0			0	x		0		282	25
D	x		0	0		x	0	x	x							x				x	0	x	x		0	x	0			188	62
E	x								x						x					x										36	73
F	0		0		0	x				x							x	x			x	x						x	x	68	68
G		x					0	x				0	0	0		0				0		0	0	0		0	0	x		279	52
H	0	x	x	0		x	0		x			x	x	x	x	x		x		0		x		x		x	0		x	282	61
I		0					x		x	x				0				x	0	x							x			196	63
J	x			x					x				0			x	0			x		x	x	0					x	112	62
K							0	x						0		0				x			x				0		x	92	57

Legend: average score 60 or above = x
average score below 60 = 0
blank means FR did no collection in that industry

percent of the 11 industries). The problem industries (primarily a cooperation problem) were beyond management's control. These are examples of the so-called Pareto distribution.

There are several factors that could influence the performance of this process. We involved the field staff directly in the development of this project to tell us what they were. The factors included the field representative, the respondent's cooperativeness, respondent burden, establishment structure, and Standard Industrial Classification industry.

We applied multiple regression analysis to this model of disaggregation quality to determine the significant contributors to variation. Of the five factors, field representative contributed 47 percent of the variation in the all industries equation which explained most of the variation in scores, more than all the other variables combined. In some of the industry specific equations, the field representative variable explained over 60 percent of the variation. All of the variables had significant t-tests at the 95 percent level or better. We could not control the variation problems associated with certain industries. We could cntrol the variation associated with field representatives.

That lead to other work to understand what we might want a field project to address by studying the results of industry analyst review of collected data. Then we focused on specific measures of field conformance in data collection (SSR) to determine the appropriate corrective action and establish control.

Initiation Review Study

Before we measured the quality of collected data we studied the Washington Office review of initiated schedules. Again we used conformance measures. We identified the types of actions taken by economists (industry analysts) in their review of the initiated establishments.

Hart and Lockerby (1982) found that most (over 75 percent) of the analysts' actions involved correcting problems created by the field staff or the processing system. The rest of their actions related to their legitimate job of "reviewing for economic reasonableness", and verifying or assigning the proper classifications. The field originated problems accounted for about 31 percent of all "unnecessary action", while the process (computer software or other systematic) problems accounted for 52 percent of the analyst' "unnecessary actions." The distribution of analyst action is shown in Table 2 below. These nine collected data elements represent about only 10 percent of the total collected data elements, yet account for nearly all the actions.

The errors are concentrated and reflect a few sources contributing the bulk of the problems. We have identified the sources for most of the "process" problems. Some of those we could correct; some we could not take action on until much later. The field originated errors required still more effort to identify the source of the errors. Now we could conduct a study of the field collection activities armed with the results from these two studies.

Structured Schedule Review

The next study takes advantage of what we learned from the two already discussed. It also takes advantage of an existing inspection process: schedule review in the regional office. The approach is again aimed at conformance; conformance to the data collection procedures. The existing review is a dependent review by regional office staff. The measuring device is a process (form and procedures) called Structured Schedule Review (SSR). Although the review is dependent, it provides much useful information. The PPIR schedules (about 6 separate forms) have enough data elements to cross-reference: data elements such as product lines, product descriptions, revenue, employment, and establishment identifiers. The cross-referencing and consistency checks are built into the SSR process so that questions can be raised to uncover problems and provide feedback.

Although the SSR process focuses on conformance, the data also provides measures useful to identify quality measurement issues. Table 3 shows that 66 percent of the errors were attributable to 25 percent of the error types. The specifications and procedures are the likely sources of error. This is another example of Juran's (1980) Pareto principle: that the bulk of the errors are attributable to a few sources. If we can identify those sources of error and correct them, then we can remove most of the error. Once the people in organization began to think in these terms, we achieved a breakthrough in results: the error rate (measured by SSR) declined by 50 percent over three months.

The last step is a very important one. The effort to establish continuous quality control and measurement is required to sustain the results. An important element in quality control is self-control. Operator errors can only occur when self-control is established. In all the processes we have discussed, some elements of self-control were missing before we studied the process. By definition, the operators (economists, field representatives, and so on) were not making any errors. All the errors were management-controllable. In order for an operator error to occur three conditions must be satisfied:

1. The operators know what they are supposed to do (have documented procedures).
2. The operators know how well they are doing (have feedback on performance).
3. The operators can change their performance if it is required (have the authority and ability to change within the process).

Now I mentioned that we saw a 50 percent reduction of errors in three months time. That occurred for two reasons. One reason was the one already discussed: the bulk of the errors were systematic. We discovered the root causes of the systematic error and took corrective action. Corrective action involved clarifying and correcting written procedures. The second reason is that we created the elements of self-control so that people who were not performing well knew it and could receive additional OJT or other instruction. Table 4 shows the nine field representatives who accounted for 42 percent of all the errors (1002/2373). In fact, just 15 our of approximately 70 field representatives accounted for two-thirds of all of the errors. Yet none of these people could be called accountable (technically) until the elements of self-control were established. Then their performance improved dramatically.

Data Capture Study

Data collection in the price programs is executed in two phases. In the first phase the establishment, outlet or household is "initiated" into the survey and the descriptive information and price quotes are collected. During repricing, the Bureau periodically collects the price data and any changes to the descriptive information. In the PPIR, the data is collected monthly (primarily) directly from the respondent on a mail shuttle form (repricing form). The process of translating that data to machine readable form on a monthly basis is called data capture.

The data capture of the monthly repricing data for index calculation is accomplished through two modes. First, all of the schedules (price quotes) are processed through optical character recognition (OCR) equipment.

Table 2

"Unnecessary"
Problems Requiring
Analyst Action

Collected (initiated) Data Element	Process Originated Problem	Field Originated Problem	Other	Total
1	4	12	34	50
2	55	18	1	74
3	52	3	1	56
4		31	1	32
5	4	-	-	4
6	-	3	-	3
7	3	-	-	3
8	-	2	-	2
9	-	2	-	2
Total	118	71	37	226
Percent	52%	31%	17%	100%

TABLE 3

ERRORS BY FIELD REPRESENTATIVES BY TYPE OF ERROR

FLD REP	# OF FMS	ERROR TYPE																																															TOT
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	
1	16	2		1					1	3				3									1				2	4	2					1					10				1						31
2	26				1	2			1	1	1			3				1	1							3			5	2							5	4			45			3		4	1				83
3	12										1	1													2	1		2	5	1				2	5	7	4	2							2			1		36
5	11			1	1	9						1	1							1						2			9	2							9		22	2	1	33	2							2				98
6	25	1	2	1	2	1	1		5	1	1	2	3	3		1		3	1	2	1	3				1	1	2	3	1					1	1	6			5	5			1			4	2		67
9	22	1	1	2	2	3			5	1				2	1						1				1	2			2		1					1		2	1	3	4			2							38
11	13	1		3	1	2	1					1		1	1						1	1		3	3			11	7	2					1					5				1		2		2		50
12	15						2		2				1	6		1									2				3	2										2	20				3						44
13	31					4	1			2	1		2	1							2				3			2	1	7		1			1	2	4	4	3	7			1				1	1		51
15	13	1		4	2			2	1	1	1	1	1	4	7					1				3	5		15	20						4		3			22	2		2			2	1		6	111
19	11				1								1	1												2		3	2	5					1	1			1	7						1	1	1		28
20	10											2	1	3										1	1			1	11	1		3					3	1		30						4	1			63
23	15	1	2	2	3	1	3		2	2			3	5			2	1			5	1	1	2	5	1		11	20	3		4		3	6	4			6	2	2		27	2	1	3		1		137
24	31	3	3	2	1	7			3	1		2	3	5		1		2	2	3	5			3	8		2	11	1	1					1	6	1	1	5							1					84
25	12	1					1		1	1	2			1					1							1		4	2								7		2									3			27
29	15			1	2	4							1	1		1									3			1	1	3						1				4	7			1				6		3	40
30	23			4	2				2	2		1		2										2	3	3	2	9	1			2		1	4		1	3		1			11			3		2		61
31	12			2			1	1		1	2			2						2				2	6		1	9	18							1	2											5		8	63
33	16					1	2				4	2									3				5		4		13	1						4	10		3	18											70
34	29		1	2	1	2		1	1	1	1	1	6	5		1			2		4			2	3	2		24	23	10				1	5	8	15	5	12	21	20	2	15	2		1	9				209
35	14				1	1								1											1	1		4	5	2																			1		17
36	12			1		3														1				1			5	4										1				3		9						28
37	10		1						2	1		1										3						1	4	5								4				2										24
38	33				2	5			1	3	4	2	12	1					2	5	4	1			4		1	17	10	12		1		1	8			7	6	35	2	2		5	6	1		2			162
42	23				3																	1			3		1	2	14									3			8		6		3	1				45
43	17		2	2		6	1		1	1	1		6	1			1			2				1	1			9	4	2																					4		45

Table 4

Error Counts by Type of Form and Field-Representative

Field-Representative identifier	Errors by Form Type A	B	C	D	E	F	Total Errors
1	9	0	3	5	18	35	70
2	30	7	4	5	40	76	162
3	9	2	0	3	7	62	83
4	13	1	0	2	11	71	98
5	25	1	0	3	40	42	111
6	9	2	0	8	28	16	63
7	22	3	4	5	59	116	209
8	23	7	1	3	8	25	69
9	24	3	5	9	35	61	137
Total							1002

As much as 75 percent of the schedules need no further processing. Then only schedules that are damaged, or any price quotes that need to be reviewed, are key entered. The data captured this way is ultimately used to update the estimation database (used for index calculation).

We studied data capture and found that the vast majority of the errors were systematic or management-controllable. In fact, out of the actual errors uncovered in the samples from the data capture process, only 10 percent were attributable to operator error. Ninety percent were due to other causes ranging from errors in computer code to the lack of clear procedures for the observed cases. Once the management-controllable problems were eliminated, the estimated error rate went to zero (0.0). Again, breakthrough was achieved. Table 5 below shows the estimated error rates for the months studied. The change in performance begins after the fourth month shown below. With this month's production, a new repricing form was introduced. Several of the systematic errors were linked to response errors caused by the old form. In this fourth month these systematic errors did not occur. However, other errors occurred including some operator error, perhaps due to processing a new form. With the fifth month's processing, these new problems were gone and so were the old problems. The seventh observation is an example of a sporadic occurence. The process was operating at a new level of performance.

Table 5

Selected Monthly Measurement	Errors Found in Sample	Estimated Erroneous Price Records in Captured Data	Estimated Price Record Error Rate
1	3	72	0.6
2	4	105	0.7
3	2	53	0.4
4	3	76	0.6
5	0	-	0.0
6	0	-	0.0
7	1	64	0.3
8	0	-	0.0
9	0	-	0.0

Conclusions

We have done many more projects like these and we continue to find the same patterns. Most errors are management-controllable. The results from these studies confirm the management principles we first discussed. The improvements would not have been achieved without an interdisciplinary approach. We found that the biggest mistake we can make is to think individuals are always responsible for most errors--before we have the processes under statistical control. All of the PPIR projects cover nearly all the areas of the Dalton matrix; however, the quality improvement job has just begun. The effort must be continuous.

For Further Study

The questions I began with are still not completely answered. This approach is just a beginning. The complexity of measuring quality from the end uses of economic data is enormous. This is the only way to determine true fitness for use. More work must be done on defining data uses and measuring quality in terms of the uses and the delivery of other services with the data. An error profile would be a major piece of that measure of quality.

However, it seems reasonable that we should know the capabilities of the processes that produce that data and remove the major sources of error. That is within our contol. The information and understanding we get from that effort may help us to better address the unanswered questions.

References

Crosby, Phillip B., (1979). Quality is Free. New York: McGraw-Hill, Inc.

Deming, W. Edwards, (1960). Sample Design in Business Research. New York: John Wiley & Sons.

, (1982). Quality, Productivity, and Competitive Position. Cambridge, Massachusetts: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Feigenbaum, A.V., (1961). Total Quality Control. New York: McGraw-Hill Book Company.

Gillespie, Stephen R., (1981). The Effectiveness of the Disaggregation Sampling Technique in the Producer Price Indexes. Masters Thesis. George Mason University.

Hart, James, J., and William Lockerby, (1982). Initiation Update Review Study. Washington, D.C. Bureau of Labor Statistics.

Juran, J.M. and Gryna, (1980). Quality Planning and Analysis. New York: McGraw-Hill Book Company.

Minton, George, (1970). "Comments on Quality Control and Research in Data Processing Programs." Arlington, Virginia: American Society for Quality Control. Conference Paper.

Shewhart, W.A., (1931). Economic Control of Quality of Manufactured Product. New York: D. Van Nostrand Company, Inc.