PERFORMANCE IS EASY TO MONITOR AND HARD TO MEASURE[1]

OGDEN R. LINDSLEY
THE LINDSLEY GROUP[2]

CONTENTS	PAGE

[1] Chapter 26 in R. Kaufman, S. Thiagarajan, & P. MacGillis (Eds.),*The Guidebook for Performance Improvement: Working with individuals and Organizations* (pp. 519-559), 1997. San Francisco, CA: Pfeiffer, Jossey-Bass.

[2] 366 North 1600 Road, Lawrence, KS 66049-9199 Tel: 785-887-6869 Fax: 785-887-6779 olindsley@aol.com.

MEASUREMENT IS THE WEAKEST LINK IN PERFORMANCE TECHNOLOGY

Measurement is the weakest link in performance technology, and in this chapter I suggest why. I list the differences between frequency monitoring and traditional measurement, and suggest using frequency monitoring to improve and validate performance in practice. I also offer examples of performance frequency monitoring at the micro, macro, and mega performance levels.

Pundits Say Performance Measurement is Crucial

Most writing about performance technology preaches performance measurement and standards. This is as it should be. For example, Gilbert's six page Foreword in Stolovitch and Keep's *Handbook of Performance Technology* tells us that an important link is to measure performance:

> Acceptable evidence about performance must rely on measurement. If science does nothing else, it measures, and we must become very good at measuring human performance. As a general rule, our clients in the workplace are not good at it, and here is where we can be of especially great help. We can have our greatest effects on human performance just by measuring performance correctly and making the information available. (Gilbert, 1992).

In his earlier classic, *Human Competence*, Gilbert devotes chapter 2, comprising 44 of his 376 pages (12%), to measurement (Gilbert, 1978).

Measurement is the third of Crosby's fourteen quality improvement steps (Crosby, 1984). Deming (1986) did not list measurement as one of his 20 points to build an effective quality program. However, his crucial data collection system is based on Statistical Process Control (SPC) (Shewhart, 1939). SPC monitors performance over time with either continuous measurements (e.g. length, weight, etc.) or counts (e.g. number of defectives, number of defects).

With this strong preaching, why is measurement the weak link? Perhaps because the pundits fail to provide leadership by example. Perhaps they fail to practice what they preach.

Many Words Few Numbers

For instance, Gilbert's above mentioned Foreword contains no sample performance measures; and few of the other performance technology pundits provide sample measurements either. There were 60 contributors to the *Handbook of Performance Technology*, yet performance measurement is not in the handbook's 610 item index, and neither are monitoring or counting (Stolovitch & Keeps, 1992). Only 3 charts and 1 table showing performance data exist in the entire 859 page volume. The 3 charts are in a chapter by one contributor, and the table is in a chapter by two other contributors. It is interesting that in his four page Afterword, Mager, in sharp contrast to the other contributors, shares performance data on one of his classes (Mager, 1992). So, all told, of the 60 contributors only 4 (7%) share performance data samples!

The type of figures used by the contributors in the *Handbook of Performance Technology* are listed in the following table with their number and portion:

Figure Type	Number	Portion
Descriptive Table	50	38%
Diagram	44	33%
Flow Chart	29	22%
Organizational Chart	1	1%
Procedure Table	3	2%
Check List	1	1%
Data Chart	3	2%
Data Table	1	1%
Total No. of Figures	132	100%

The handbook contributors primarily used descriptive tables, diagrams, and flow charts. These "what you should do" figures make up 90% of the figures in the handbook. Many diagrams and flow charts might be appropriate for a theoretical handbook on general systems theory, but it is sad for a field whose experts preach measurement as an essential component!

Journal Authors also Write Many Words Few Numbers

Further proof of the "many words, few numbers," tradition in performance technology is demonstrated in *Performance and Instruction*, one of two official journals of the National Society for Performance and Instruction. A frequency count of the number of different types of figures contained in volume 28, 1989 is shown in the following table:

Figure Type	Number	Portion
Descriptive Table	56	20%
Diagram	169	61%
Organizational Chart	1	0%
Procedure Table	18	7%
Check List	27	10%
Data Chart	5	2%
Total Number of Figures	276	100%

Note that, as in the *Handbook of Performance Technology*, by far the most common *Performance and Instruction* figures are descriptive tables, diagrams, and flow charts. Here again, these "what you should do" figures make up 88% of volume 28, 1989. The most common type of table is not a data table, but a procedure table. Practicing performance technologists tell each other how to do things, but seldom report any quantitative results.

Only a Few Give us Numbers With their Words

Along with Bob Mager's Afterword, Aubrey Daniels' *Performance Management* is a closer approximation of displaying performance measurement for the reader (Daniels, 1989). Both "counting" and "measurement" appear in its 120 item index. Daniels points out that counting is the preferred performance measure because it is the simplest, easiest and most reliable. However, he falls short of setting counting or charting standards. Four of the 16 chapters concern performance measurement ("pinpointing," "measurement," "the right pinpoint," and "feedback in graphs"). These chapters comprise 67 (26%) of *Performance Management* 's 259 pages. The first 2 of Daniels' 6 performance improvement steps are "pinpoint" and "measure." The book contains 42 figures displaying samples of performance. However, 21 of these are performance percentages and the other 21 are only stylized (made up) curves. There is no doubt that real data would have been better, yet even stylized curves help the reader a bit more than nothing.

The above examples clearly demonstrate that measurement is the weakest link in performance technology - so weak it is almost non-existent. So what accounts for the fact that we talk measurement and don't do it? The reasons for this are historical, cultural, but most of all practical. There is nothing inherent in performance that makes counting it difficult. In fact, both behavior and accomplishments are easy to count and time. Even private, inner behaviors such as feelings, attitudes and urges are easy to count by the person having them, and the period over which they were counted is easy to record (Duncan, 1971). These counts per minute, per day, per week, per month, or per year give us frequencies which are the most sensitive and universal performance measures available.

WHY MEASUREMENT IS SO WEAK

Historical Reasons Why Measurement Is So Weak

There are five historical reasons why performance measurement is weak.

First, **performance technology was influenced by Programmed Instruction** in the late 1950's and 1960's. Even though Skinner called frequency (rate of response) his most important contribution (Evans, 1968), he dropped both rate of response as a datum and high frequencies as goals when he turned from his free operant laboratory to develop Programmed Instruction. The programmed learning performances were timed but not reported as frequencies. For example, the results of using one of the first programmed texts (Holland & Skinner, 1961) in Skinner's undergraduate course (Natural Science 114) were reported by Skinner as follows:

> On the average, a student went through the set in about 14 1/2 hours. The easiest disk took 8 minutes, with a range from 5 to 13 minutes, and the hardest took 28 minutes, with a range of 17 to 80 (Skinner, 1983, p.139).

Holland or Skinner could easily have computed frequencies on these data. Since there were 29 frames on each disk, the overall rate of performance was 1.7 frames per minute. The easiest disk was 3.6 frames per minute with a range of 5.8 to 2.2 and the hardest disk was 1.0 per minute with a range of 1.7 to 0.4 frames per minute. But they didn't! Almost everyone else in Programmed Instruction followed Skinner's lead and abandoned frequency.

Second was the **influence of Gilbert's book *Human Competence*** (1978) throughout the late 1970's and 1980's. In this bible of the field Gilbert defined and named performance technology, clearly separating it from the earlier Programmed Instruction. He set standards for performance technology, but set no standards for performance technology measurement. This was uncharacteristic. Years earlier, in 1957, Tom Gilbert had worked on systems for standard performance measurement in my Harvard Medical School human behavior laboratory at Metropolitan State Hospital, Waltham Massachusetts. In the intervening years he apparently lost what had been a strong interest in standard measurement. It had been the bond between us, the thing we shared, and the reason I invited him to use one of my laboratory rooms. Gilbert's book stressed the requirement for measuring accomplishments, but he listed 3 classes, 9 dimensions and 19 performance measures with no guidelines for choosing between them! He opened a Pandora's box of different and unstandardized measures (Gilbert, 1978, pp. 43-50).

Third was the **influx of trainers into performance technology during the 1970's and 1980's.** These trainers came from traditional human resource or educational research departments and were schooled in using 5 and 10 point rating scales (Thurstone & Chave, 1929; Likert, 1932) for measurement. Their method of analysis was the analysis of variance and most measurement was summative, occurring after the performance intervention to determine its effect. These statistically schooled trainers so strongly believed that you must transform data before you can analyze it that they called original data "raw." Implying that raw data could make you ill, it followed that their methods should be used to "cook" all data before ingestion. The statistical methods were taught by rote memorizing of formulas, by which the learners gained little real understanding. Therefore, we have a generation of performance technologists who believe we must use statistics, but who both poorly understand and frequently fear statistics. As a result, most avoid measurement entirely, unless they yield to a quick, nasty 5 point rating scale.

The fourth reason is that **measurement experts caution** their readers about the complexity of performance measurement and urge industrialists to hire statisticians to handle this complexity. In Gitlow & Gitlow (1987) chapter 16 is " How to hire a statistician" and lists 11 requirements of an effective industrial statistician along with 277 university statistics departments which train industrial statisticians. How better to warn practitioners that they better not measure performance by themselves!

The fifth and final historical reason is that **measurement experts overwhelm** their performance technology readers with too many options and too few guidelines. Chapter 9 of the Handbook of Human Performance Technology (Geis & Smith, 1992) covers the when, what and who of evaluation. Then chapter 10 (Smith & Geis, 1992) goes on to discuss how to measure performance, but gives few guidelines. Smith & Geis include a summary table of 18 data-collection techniques used in 331 evaluation studies to measure 10 different performance dimensions with no further description or ranking of techniques. How better to confuse and scare off practitioners?

Cultural Barrier to Performance Measurement

Culturally, the barrier to performance measurement is even more debilitating. For over 6,000 years we have looked at the world through our fingers. We instinctively use "add scales," in which increments are of 5 (one hand), 10 (two hands), or 100 (percents or twenty hands) are added. The problem is that the world around us is a multiply world (Meadows, et al 1972, p.25). Everything in it grows by multiplying and decays by dividing. In distributions, performance frequencies spread by equal multiples up and down. If the middle performer produces 20 per hour, and the bottom performer produces 10 per hour, the top performer should produce 30 per hour. Wrong! The top performer will produce 40 per hour! The down-spread is 20/10 or x2 and the up-spread is 20 x2 = 40. When we look at our multiply world through base ten add scales we get a distorted view. We try to compensate for this distorted view, not by using standard multiply scales, but rather by using a large assortment of add scales which both complicate and distort the picture.

Two commonly used scales illustrate this cultural problem of looking at multiplying things through add scales.

Decibel Sound Intensity Scale

The first illustration is the scale of intensity used in sound research. The intensity of sounds around us multiplies in sound pressure energy. A whisper is 10 decibels, and the noise on the flight deck of a flying B-24 WW II heavy bomber was 120 decibels. Now the question, what is a decibel? How many times more intense is the B-24 flight deck over a whisper? Few of us really know.

Historically, to handle the multiplying sound intensity, a times 10 increase in sound pressure was set at one bel (after Alexander Graham Bell). The bel was subdivided into ten smaller steps which were called decibels but were add units on the basic multiply scale. Now the confusion – halfway up from one bel (10 decibels) to two bels (20 decibels) is 15 decibels. If 10 decibels represents x10 and 20 decibels represents x100, does 15 decibels represent x50 as you would expect? No! 15 decibels is x31.6, (the square root of 10 = 3.16) a point half-way up the multiply scale from 10 to 100!

Our B-24 flight deck sound intensity was 120 decibels, or 110 decibels greater than a whisper. Does this mean the B-24 is 110 times more intense than a whisper? No! It is x10 eleven times in succession or 10 billion times more intense! It would have been clearer to have left the intensity in actual original values. It would have been much clearer to have left the multiply scale alone and not put add values on it. It would have been clearer to say the B-24 flight deck had the sound intensity 10 billion times greater than a whisper. When we look at multiplying things through an add scale we are easily confused.

Richter Scale of Earthquake Intensity

The second illustration is the Richter scale of destruction produced by earthquakes. As earthquakes increase in size their destructive force multiplies. Rather than simply reporting the force, the quake is rated on a Richter scale. Most of us think that an earthquake of Richter 8 is as much more destructive than a 7 as a 7 is of a 6. Wrong! Those of us who have used logarithms think in multiples to base 10 (logarithms) and that a Richter 8 is 10 times more destructive than a Richter 7 which is 10 times more destructive than a Richter 6. By this logic, a Richter 8 is 10 x 10 or 100 times more destructive than a Richter 6 quake. Wrong! The Richter scale is add numbers on multiplying destructive forces - but not to base 10 multiply! It was not based on the amount of force, but rather on the amount of displacement of the needle on a standard seismograph. So actually a Richter 8 is about 35 times more destructive than a 7, and a 7 is 39 times more destructive than a 6 (Gere & Shah, 1984, p.79). Therefore, a Richter 8 is 35 x 39, or 1,365 times more destructive than a Richter 6. This is 13 times more destructive than the base ten multiply logic of times 100 assumed! Once again, when we look at multiplying things through an add scale we are easily confused.

There are many other illustrations of the confusions produced by attempting to simplify multiplying phenomena by putting them on a false add scale: tornado intensity, hurricane intensity, hotness of pepper, and severity of an insect sting. It is evident that scale confusion is a cultural problem and definitely not limited to performance technologists.

Practical Reasons Why Traditional Measurement Cannot be Used

There are two overbearing practical reasons why traditional measurement techniques cannot be used easily by performance technologists.

The first practical reason is, simply put, **clients don't like it!** It is too cumbersome, and most customers react negatively to any suggestions of traditional measurement procedures. If you force clients to use them, you may get the data, but you will lose your client! About the only measurements that performance technologists can use in practice are the measurements that their clients had in place before they hired the performance technologist. These are usually reported as percentages and therefore are not very useful.

The second practical reason is, **performance technologists are not paid for their effectiveness.** We are paid by the hour, and we are paid to please our clients. If we were paid some portion of the money we saved our clients over the next five quarters, we would work harder to put measurement systems in place. We would work harder to convince our clients that a monitoring system which they might at first dislike, will in the longer run make their business more profitable. We would work harder to develop comfortable, easy, non-threatening, monitoring systems. But, since we are paid primarily for the smiles we produce, we play games with our clients and ask them only if they are happy.

I have previously reported similar practical reasons why effective teaching techniques are not widely adopted by schools (Lindsley, 1992a). Simply put, teachers don't like them, teachers are paid by the hour rather than their pupils' gain, and parents don't demand effective teaching.

OUR SOLUTION - MONITORING PERFORMANCE

So, in face of these negative observations, what solutions are at hand? How can performance - both behavior and accomplishments be easily monitored at work? Although counting is common in the work place very little is written about it. Monitoring by counting is not seen as a substitute for traditional measurement. In fact it is ruled out by implication - it's not cooked.

There are measurement experts that say you must objectively define a thing before you can count it. Wrong! You can even count unknowns. You can keep track of the time and chart the frequency of unknown things you encounter each day. The daily frequency of unknowns is very high when you are in a foreign place, and very low when you are in a familiar place.

Other measurement experts say you cannot count a mix of apples and oranges. Wrong! You can easily count a mix of apples and oranges - what you are counting is fresh fruit. If you had counted oranges, lemons and limes, you would have counted citrus fruits.

Skinner's Performers Monitored Their Own Performance Frequency

Although seldom identified with the field of measurement, B. F. Skinner made major measurement contributions (Lindsley, 1992c). His contributions are so novel, so major, and so revolutionary, that most measurement experts don't see them as measurement. Skinner gave us self-recorded frequency (which he called rate of response). In his classic *The Behavior of Organisms* he wrote:

> All the curves given in this book (except those obtained by averaging or those extending over a number of days) are photographic reproductions of records made directly by the rats themselves (Skinner, 1938, p 60).

Skinner Monitored With Standard Slope Charts

This frequency of responding was automatically displayed on standard charts on which the data made slopes which were represented by standard grids of 1, 2, 4, and 8 responses per minute. These standard slope charts were called cumulative response records and were produced by the rats through the automatic recording equipment. The complexity of measurement was done automatically by the recording system, not by the performers. The performers merely pressed a lever. The calculation of the frequency (number per minute) was done by the automatic recorder.

Skinner thought that frequency and the self-recorded, standard slope chart were so important that when asked what his most important contributions were he said:

> "My major contributions are rate of response and the cumulative response recorder." (Evans, 1968).

Self-Frequency Monitoring Effective in Precision Teaching

In the first systematic application of laboratory free-operant conditioning methods to humans, I restricted the measurement system to frequencies of patient's performance and symptoms. These frequencies were automatically recorded on standard slope cumulative recorders (Lindsley, 1956), just as in the rat and pigeon laboratories. When others started applying free-operant methods in more open settings such as school classrooms, they dropped both self-recorded frequency and standard slope records. They substituted in their place observer recorded percent-time-on-task or percent-correct and a wide range of non-standard stretch-to-fill charts (Bijou & Baer, 1961).[3]

Agreeing with Skinner, that his major contributions were self-recorded frequency and standard slope recording, I closed my laboratory and went into teacher training at the University of Kansas to secure Skinner's contributions within public school education.

My students and I successfully developed self-recorded frequency monitoring by both regular and special education public school students from pre-school through high school. This has been called Precision Teaching and has its own journal in its tenth volume and annual international conferences in their eleventh year. Several laws of performance and techniques which can double performance each week have been discovered over the past 25 years (Lindsley, 1992b).

Standard Celeration Chart Developed to Simplify Charting by Children

Skinner's standard chart had cumulative responses up the left and its standard slope was frequency (performance). In Precision Teaching we developed a chart with frequency up the left to simplify chart making and reading. Its standard slope was acceleration in frequency (learning). Pre-schoolers were taught to chart their own performance on this standard celeration chart. The daily chart covered the full range of human performance frequencies from 1 per day through 1 per minute up to 1,000 per minute (1,000,000 per day). The chart was designed so that a line from the lower left corner to the upper right corner (34 degrees) was doubling in performance every week. The chart was also designed to cover 140 days or one school semester of 20 weeks (Pennypacker, Koenig & Lindsley, 1972). A family of charts with the same angle slopes were made for daily performance (doubling every week), weekly performance (doubling every month), monthly performance (doubling every six months, and yearly performance (doubling every five years)[4] .

Efficient Industrial Monitoring Requires Standard Charts

The values of using common performance charts across a company has been known for over 30 years, even though few companies use them today. Ralph Cordiner of General Electric wrote that a common monitoring system would:

1. Permit all managers to record and plan their own performance.
2. Permit each manager to detect a deviation in time to do something about it.
3. Provide worker appraisal, selection, and compensation based on performance.
4. Motivate workers by their records of their own effectiveness.
5. Simplify business communications with common quantitative concepts and language.
6. Permit executives to evaluate performance in 100 different businesses without becoming involved in the operational details of each (Cordiner, 1956, pp. 95-98).

---

[3] "Stretch-to-fill" is a name I have given to the type of charts we were taught to make in school, and those that most computer graphics programs make. Figure 1 is a stretch-to-fill chart. They must always be made after the data have been collected. The vertical or magnitude scale of the chart is made to encompass the range of the performance magnitudes collected. If the lowest number is 60 and the highest number is 180, then the scale is stretched so the data fit neatly inside - perhaps a scale from 50 at the bottom to 200 at the top. The horizontal or time scale is also stretched to fit the data so if the first of five observations were collected on May 8 and the last were collected on October 3, the horizontal scale would be stretched from May 1 to October 31 to include the observations. Even worse, real time might be thrown away and the time dimension completely distorted to read from 1 to 5 equally spaced observations as was done in Figure 1. If stretch-to-fill charts were used in animal pictures, an elephant would look like a long-nosed, hairless mouse! They both would appear the same size because they had been stretched to fill the same size rectangle.

[4] A full complement of Standard Celeration Charts, chart transparencies, counters and timers are available by mail order from Behavior Research Company, Box 3351, Kansas City, KS 66103.

Even Rensis Likert, known for his attitude scales, recognized the importance of common measures used across a company (Likert, 1958).

Charting Quality Pairs Guarantees Both Quality And Fluency

From their beginnings in 1965 precision teachers always viewed their students' charts of both correct frequency and error frequency. We compared frequency with percent correct in our first year of classroom research. Classroom frequencies recorded 40 times more effects of curricular changes than did percent correct on the same practice sheets with the same children, in the same rooms on the same days (Holzschuh & Dobbs, 1966).

In our first monitoring of outer and inner personal and social behavior it was also clear that positive and negative outer behaviors independently accelerated and decelerated from each other. Positive and negative inner behaviors also independently accelerated or decelerated (Duncan, 1971). At the time we called these "fair pairs," thinking it only fair to replace a negative behavior with a positive behavior.

Later the full impact of the independence of correct and error learning became clear. If correct and errors are reciprocally related as most of us assume, then there would be only three major learning patterns: 1) correct increasing and errors decreasing, 2) corrects maintaining and errors maintaining, and 3) corrects decreasing and errors increasing. In practice in only one classroom with 29 children, eleven different learning pictures were recognized by the teacher and students (All, 1977). This demonstrated the independence of correct learning from error learning and proved we must chart both independently.

In application to industrial performance the principle of independence has also held true. For this reason we must always count and chart an acceleration target/deceleration target pair to guarantee both quality and fluency.

Fluency Aims Successful in Education and Business

Fluency, one of Precision Teaching's most powerful techniques, was discovered by Eric and Elizabeth and their students (Haughton, 1974). They found that when you practice performance far beyond full accuracy up to very high frequencies (for example: basic add facts to 300 digits per minute) You get more retention, more endurance, and more generalization to other workplace situations (Haughton, 1981).

Self-recorded and self-charted performance frequencies have been combined with Direct Instruction and Tiemann-Markle instructional design to produce really powerful learning. Students at Morningside Academy in Seattle are given a money-back tuition guarantee if they do not gain over two grade levels per year. In seven years Morningside has never had to refund (Johnson, 1989). With an adult literacy program for the Job Training and Partnership Act, Morningside agreed to be paid only for students who progressed two grade levels in two skills in 21 months. Twenty nine of the 32 students exited with skills above the eighth-grade standard. Their attendance was 3.8 days per week; they spent one hour in each of two skills per day; and they gained an average of 1.7 grade levels per month. This is 10 times the gain required by the U.S. government standard (Johnson & Layng, 1992).

When company product knowledge performance is practiced beyond accuracy to high fluency the performers have not only higher retention, endurance, and application, but also develop high self-confidence (Binder, 1990). Precision Teaching to high fluency aims has been successfully applied in sales training for new product knowledge in banking, computer software, and bio-medical companies (Binder & Bloom, 1989).

Limitations of Percent

Percent is a dangerous measure. It is fine to use as a standard in comparing 2 or more static portions. For this reason I head columns of percentages in tables "Portion" to make this clear. But percent is very misleading when used to monitor the change in magnitude or portion in a time series. Percent change is not symmetrical. For example, many of us do not realize that if you add 20% and then subtract 20%, that you are not back to where you started from; you are actually below where you started. In fact, if you start at 100 and add 20% and then subtract 20% ten times in a row, you will be down to 66.6.

A second problem is that percents can be very misleading in monitoring performance. For example, in using percent of calls closed to monitor life insurance sales, the salesman with the highest percent sales, often would sell a policy on his first call, then go play golf for the rest of the day because he had just made 100% sales.

A third problem is that percent overlooks the absolute values and frequencies of performance. So the whole dimension of fluency, or speed of responding, which is important in product knowledge and other business skills, is totally lacking.

Percent is expensive, because it takes time to calculate and it produces errors. There are other problems with percent too numerous to detail here.

Skinner knew of these limitations of percent when he wrote the following advice to young psychologists overcome with the expanding research literature of the late 1960's:

> Some principle of selection is needed, and a useful guide is the significance of the variables studied. A glimpse of the coordinates of the graphs in an article will usually suffice. A good rule of thumb is as follows: do not spend much time on articles in which changes in behavior are followed from trial to trial, or in which graphs show changes in the time or number of errors required to reach a criterion, or in amount remembered, or in percent of correct choices made, or which report scores, raw or standard (Skinner, 1969, p.93).

Exemplary and Lemony Days

Gilbert (1978) introduced the concept and term "exemplar" to describe the best performers on a particular job in the workplace. The performance of these exemplars should be observed to discover the very small differences between how they did their jobs, and how the other less productive workers performed. In our work in Precision Teaching with handicapped persons in sheltered workshops our attention was drawn to significantly better performance on certain days. In 1971 we called these "exceptional days." Examination of the work conditions on these exceptional days resulted in improving the performance on all the days. I later realized that we should extend Gilbert's term from exemplary workers to the exemplary days of a single worker.

In 1966 in our work with parents of disturbed children we noticed significantly worse days with much higher frequencies of symptoms. We called these "exceptional days" also, and analyzed their conditions to improve the children's performance. In a workshop at the 1992 International Precision Teaching Conference I named these "lemony days." In a performance distribution of workers, a significantly lower worker, whose performance should be examined for barrier conditions, would be called a "lemon."

In a performance frequency distribution of different workers, the best performer would be an "exemplar" and the worst a "lemon." In a time series performance frequency chart a single performer might have unusually good "exemplary days" or unusually poor "lemony days." The words "exemplar" with "exemplary," and "lemon" with "lemony" are in the unabridged dictionary with their correct meanings (Gove, 1961).

DIFFERENCES BETWEEN MONITORING AND MEASURING

What most of us know, dislike, and fear about measurement comes from our unpleasant experiences with hypothesis testing and normal curve statistics in university courses on measurement. We have been preached at to hypothesize, test, and measure. We have not been taught how to count and monitor. To clarify some of the differences between monitoring and measuring here are the five "W's" of each:

5 W's	MONITORING	MEASURING
Who	Self	Third Party
What	Frequency, Quality	Gain
Where	Workplace	Personnel Center
When	Continuously	Periodic
Why	Improve, Do your best	Select workers, Methods

Features of Monitoring Compared with Measuring

The following features will help to further distinguish between monitoring and measuring:

FEATURE	MONITORING	MEASURING
Data	Original	Derived (%, Mean, Range, St.Dev.)
Detail	Full	Abstractions
Appearance	Attractive	Official
Assesses	Pace, Quality, and Rate and Type of Change	Quality only
Time Dimension	Continuous in Real Time	Brief Snapshot
Performance	Constructed	Multiple Choice
Forecasting	Straight line on Standard Chart	Not Possible
Format	Developed as progress	Pre-packaged
Feedback	Always Continuous	Usually None
Relates to	Self	Peer Normed Group
Failure	No Failure - Information Only	Fail or Pass Criterion
Validation	Easy - at work station	Difficult
Approach	Coach	Umpire or Judge
Influence	Profits	Prophets

Statistics used In Monitoring Compared with Measuring

The statistics used in monitoring are very different from those we were taught to use in measuring. Monitoring statistics always refer to original counts taken directly from the monitoring charts. There is not space here to go into these differences in detail. However, it seems important to list the statistical dimensions and methods used in monitoring compared with those used in traditional measuring in the following table:

STATISTIC	MONITORING	MEASURING
Type	Descriptive	Evaluative
Parameters Assumed	Non parametric	Parametric
Evaluation	Formative	Summative
Determined By	Read Standard Chart	Mathematical Calculation
Measurement scale	Base 10 Multiply	Base 10 Add
Central Tendency	Middle	Mean
Daily Variance	Daily Bounce	Standard Deviation
Distribution Dispersion	Spread on chart Highest divided by Lowest	Range Highest minus Lowest
Outlier Probability	No. of Bounces Away From Rest of Values	No. of Standard Deviations Away From Mean
Research Design	Time Series	Small Group
Subjects Required	Single Subject	Five to Ten per Group
Experimental Controls	Each a Phase in Series	Control Group
Accuracy	Both Corrects and Errors	Percent calculated
Frequency	Number per minute	Not available
Acceleration	No. per min. per week	Not available

Examples of Monitoring Compared with Measuring

Examples of the differences between monitoring and measuring pinpoints often help clarify their basic differences. The following table lists pinpoint differences between monitoring and measuring for eight different business applications.

Field	Monitoring	Measuring
Air Flight	Air Speed, Altitude, Turn & Bank, Fuel, Manifold Press., etc.	Percent of flights that arrived on time.
Real Estate Sales	Sale closings and Unsuccessful Calls Counted and charted each Day.	Percent of Calls that produced sales.
Telephone calls to service desk	Calls successfully and immediately serviced, and Calls not immediately serviced charted each day.	Calls rated on a scale from 1 to 10 in degree of problem and averaged each day.
Veterinary Practice	Number of animals treated, and number of treated animals that died charted each day.	Percent of Customers satisfied each day.
Retail Hardware Store	No. of: Customers purchasing what they came for; Customers purchasing extras; Customers who didn't find what they came for. Counted and charted each day.	Percent of customers that left without purchase.
Graphic Artist Logo Designer	Presentable Logo drafts and unpresentable Logo drafts Counted and charted each day.	Presentable Logo drafts rated on suitability scale from 1 to 10 each day.
Hotel Guest Service	Guests given "thankyous" and "stops" at check in. Employees count, chart and turn in ones they receive from guests each day. Guests get discount for "thankyous" and "stops" used at check out from hotel, by turning in stubs used.	Guests asked to fill out service survey.
Safety Program Behavior	Workers given "thankyous" and "stops" to hand out to other workers for safe acts and unsafe acts. Stubs turned in to supervisors Workers chart safes and unsafes each day.	Percent of acts that are safe.
Safety Program Accomplishment	Near misses, accidents, lost time accidents, and deaths Counted and charted each day	Percent of accidents that are lost time.

Monitoring Always Measures, Measuring Never Monitors

It is interesting to note that in monitoring you get the best of both worlds, for your monitoring data can always be used to measure the effects of your program. However, if your only data are measurements, then these cannot be used to improve your program in progress.

Back Counting from Company and Personal Records

One of the nicest things about monitoring from counts, is that often you can get baseline counts for previous years from client's ordering books, service records, salespersons' appointment books, diaries and telephone logs. It is relatively inexpensive to "back count" this information and enter it into calendar synchronized charts. These charts provide the baseline information necessary to demonstrate the results of your performance improvement program.

Most importantly, a couple of years of baseline charts often reveal regular rhythms in employees performance (e.g. safety) or clients performance (e.g. sales). Precise forecasting of these rhythms which usually will not yield to statistical Fourier analysis can easily be made graphically on charts. Forecasting performance rhythms is necessary to separate them from effects of your improvement program. The sales rhythms for one product (e.g. tires) are usually independent and totally different from the sales rhythms of other products (e.g. batteries) even with the same retailer, same store, and same customers. Accurate forecasting of sales rhythms can also produce real savings in interest on the inventory that must be carried.

Personal appointment books and telephone logs can be used to back count and back chart micro level performance baselines on employees and small groups. There is a storehouse of information in those cardboard boxes in the back closets of most corporations.

Dead-man and Leave-it Tests for Performance

There are two simple, practical tests for whether you have performance or not. The dead-man test for behavior and the leave-it test for accomplishment

The dead-man test for behavior was developed by me in workshops for parents of retarded children in 1965. Simply put, if a dead man can do it, it isn't behavior and you shouldn't waste your client's money trying to produce it. For example: accident-free days do not pass the dead-man test. The dead never have accidents!

The leave-it test for accomplishment was developed by Tom Gilbert in 1962 in workshops on behavioral engineering. Simply put, if you can leave it behind at the plant when you go home at the end of the day, it is an accomplishment. For example: Increased awareness of safety is not an accomplishment, for you take it with you when you go home at night.

I have previously described these tests in a little more detail (Lindsley, 1991). This brief description should give you the ideas. These are great tests; they'll help your pinpointing a lot!

MICRO, MACRO, AND MEGA LEVEL MONITORING

Counting and charting can be used effectively to monitor all three levels of organizational contribution (Kaufman, 1992). Self-monitoring can help improve micro level contributions to individuals and small groups within the organization. Self-monitoring can help improve macro level products that the organization delivers to its external clients. Regular mega level monitoring can help an organization plan what it delivers to society at large. Examples of monitoring on standard celeration charts at these three levels of organizational contribution along with an example of input level monitoring follow.

As mentioned in the introduction to this handbook, micro level monitoring examples are much easier to find than macro and mega level examples.

Example of Micro Level Monitoring - Company Safe and Unsafe Acts

The effects of a company sponsored safety program on the safety of chemical handlers is an example of performance technology applied to benefit a group of workers within the company. Therefore this is micro level monitoring.

After an effective safety program was put in place, the supervisory safety team observed work places and recorded the number of both safe and unsafe acts. It would have been better if we had a record of the frequency of safe and unsafe acts before the safety program, but this was not available, and the client would not delay the program to collect a baseline for use in later determining program effectiveness.

Figure 1 shows the chart the client made of his observed safe acts[5]. The client's safety team converted the actual counts of safe and unsafe acts to the percentage of safe acts, thinking this would act as a reward and motivator to the workers since it showed such nice high percentages of safety. They thus broke the rule to chart original frequencies[6]. Also, because the safety team was a little sensitive about not having made their observations regularly, they hid that fact by charting the time as order of observation rather than real time. They thus broke the rule to chart in real time. Obviously they also broke the rule to use a multiply scale and standard charts.

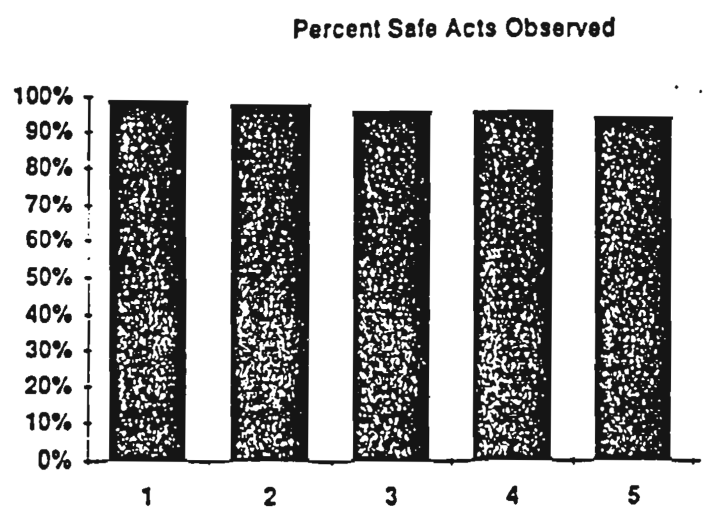


Figure 1. Percentage of safe acts observed in each of five supervisory observations of chemical handlers at work.

[5] The source of these data is not cited to honor the wishes of the client.
[6] The 11 step job aid at the end of this chapter can serve as a list of performance monitoring rules.

Figure 2 shows the original frequencies of safe and unsafe acts from which Figure 1 was derived, charted in real time on a common standard celeration chart. Note that the frequency of unsafe acts is accelerating at a factor of x1.4 per month. You can tell the acceleration by comparing the slope of the unsafe acts with the celeration guide with its fan of numbered slopes at the right center of Figure 2[7]. Note that the unsafe line is parallel to the 1.4 line on the guide. Increasing by 40% each month! The chart calls for rapid action to avert this acceleration. This worsening trend is not as apparent in the percent chart in Figure 1. Neither can this acceleration in unsafe acts be quantified in Figure 1.

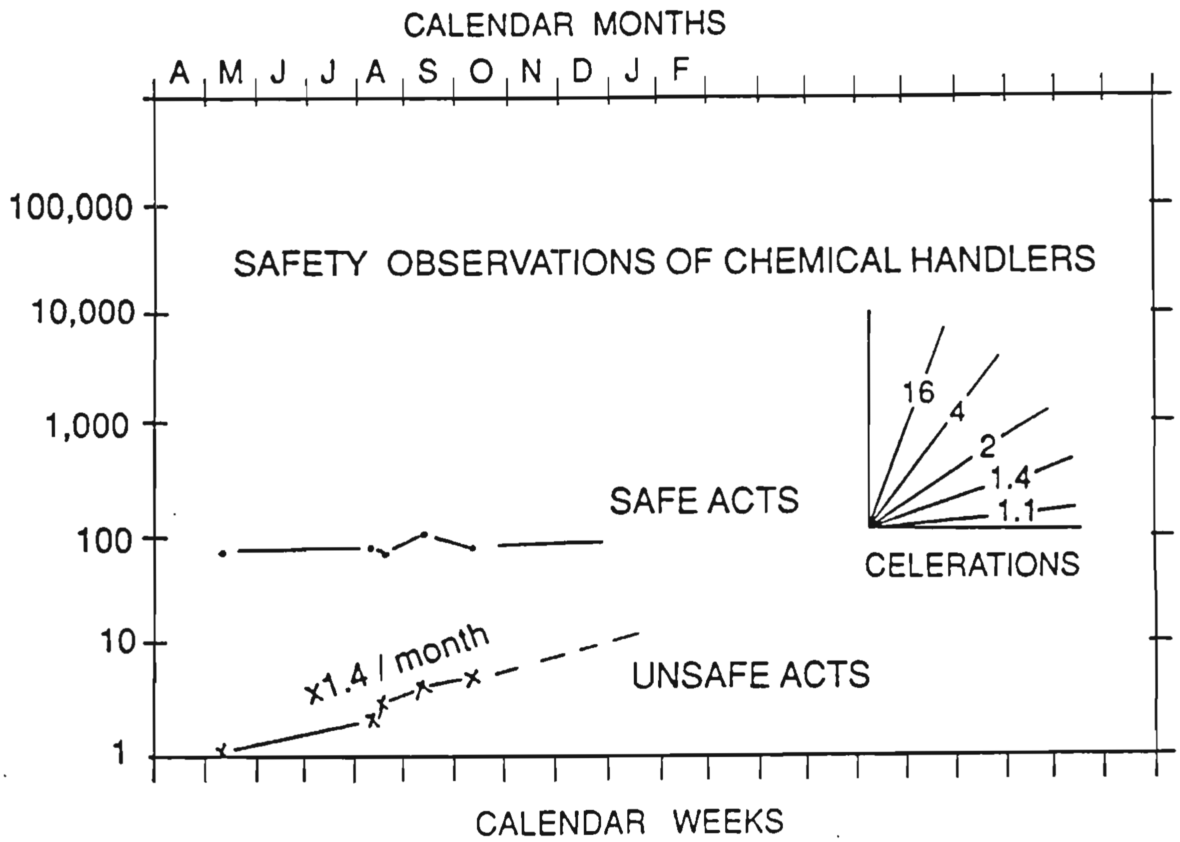


Figure 2. Original data from Figure 1 charted on standard celeration chart. Number of safe and number of unsafe acts per day made by chemical handlers on each of five observation days charted in the calendar week in which each observation was made.

Charts of original frequencies should be used to inform and improve performance. Charts should not be distorted in order to reward or please performers. Charts should be designed to inform, not to reward. Significant accelerations in low frequencies are important to see. Remember it takes only one unsafe act to produce a death.

---

[7] The celeration guide numbers slopes at x1.1, x1.4, x2, x4, and x16 per celeration period. Once you are familiar with the standard celeration chart these guidelines are no longer necessary because you have learned to read celerations directly. Celeration guides appear on the three other standard celeration charts that follow in this chapter.

Example of Input Monitoring - Waiting Time in Hospital Admission

One of the major objections that patients have of large urban hospitals is the long waiting time in admissions. A large, midwestern, university hospital took 2.3 hours to admit patients in July 1987. In July 1988 the hospital's quality improvement team put one person in charge of all admission tasks for each patient, and made other administrative cut-backs.

Figure 3 shows that the admission time reduction program divided the admission time by 6 from 120 to 20 minutes in its first 3 months.

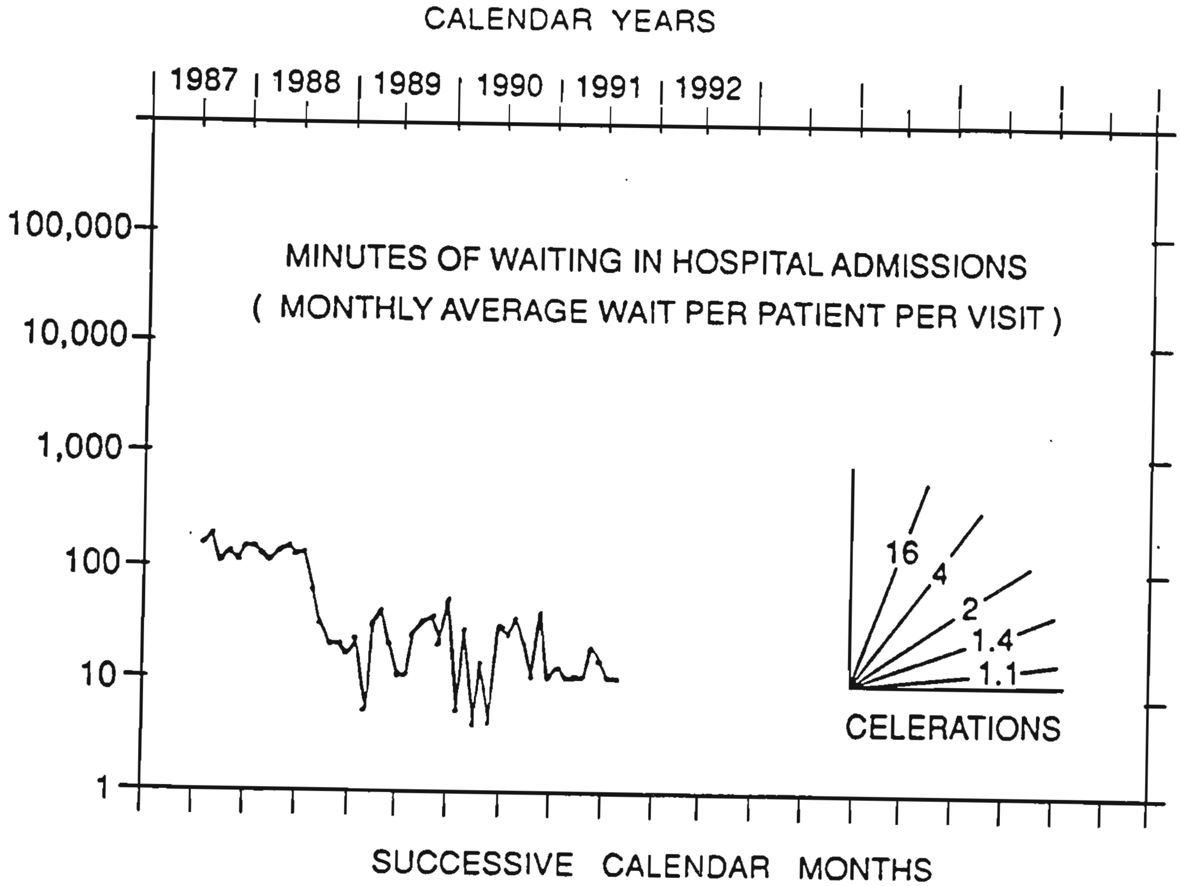


Figure 3. Monthly average wait per patient per visit in admissions of a large, university affiliated, midwestern, hospital charted on a standard celeration chart in successive calendar months,.

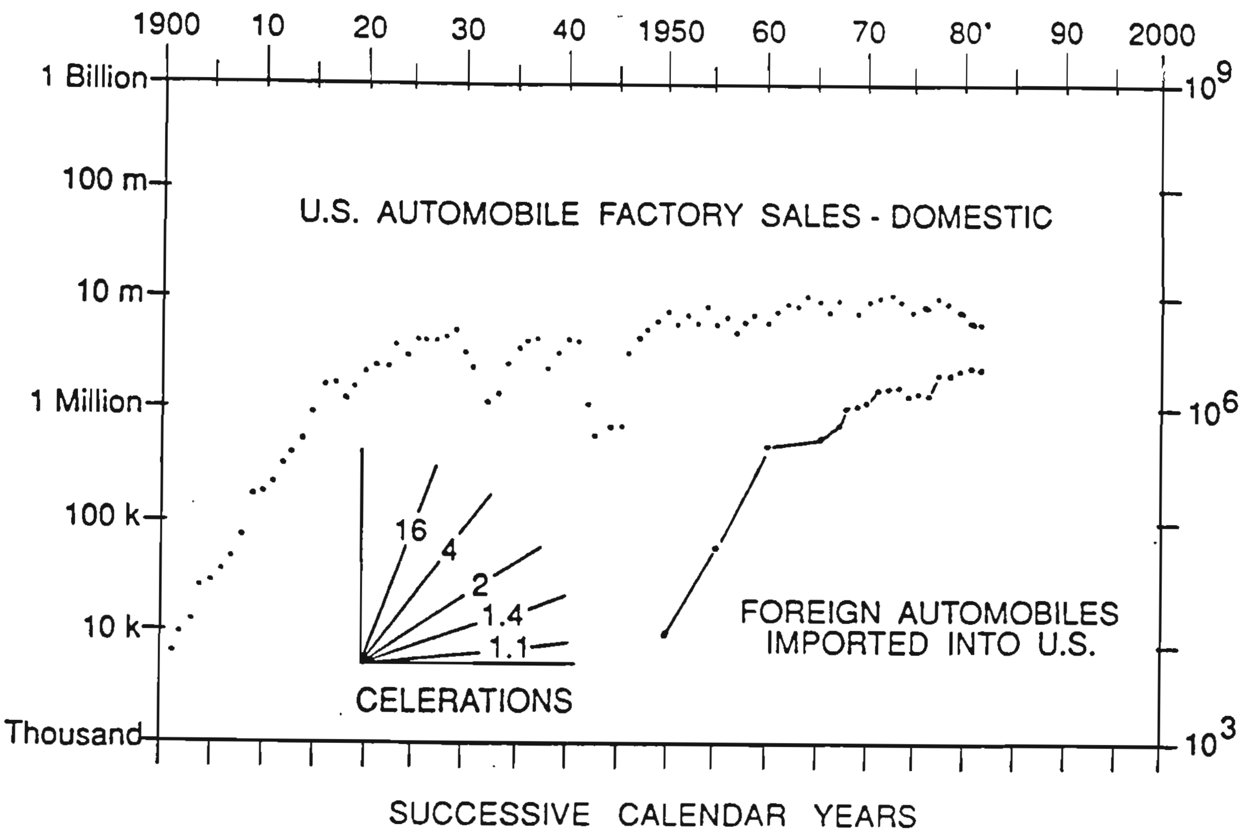
After this immediate deceleration, the waiting time maintained with a bounce from 4 to 50 minutes with a middle at 20 minutes for the next 24 months. The hospital quality team was not monitoring on charts with a multiply scale so it was hard for them to see the variations around low values as we can see in Figure 3. Had they used standard celeration charts and knew how to read them, they would have seen the tendency to level off at 10 minutes, which was set as a goal in the quality time reduction program. Deming (1986) is very much against quality goals for this reason.

The quality team also would have seen the exemplary months of February 1989 and January 1990 with 5 minutes, and March and May of 1990 with only 4 minutes waiting time. These exemplary months show clearly that since waiting times as low as 4 minutes occasionally occur, there is no reason they shouldn't be produced all the time. The exemplary months should have been examined for causal factors. Had they been, and action taken, the waiting time could have been reduced certainly to 4 minutes and probably to 1 or 2 minutes.

Example of Macro Level Monitoring - US Versus Foreign Made Cars

Perhaps the greatest economic disaster of the late twentieth century is the demise of the U.S. automotive industrial giants. It is commonplace to remark that the Japanese manufacturers caught U.S. manufacturers asleep, and took over their market. What is not commonplace is knowing what Detroit could have been doing to prevent this market take over. Had U.S. manufacturers merely charted their combined annual domestic production against all foreign imports on a multiply scale, they would have seen this coming by 1970 and had at least 10 years to manufacture smaller, more reliable domestic cars.

Figure 4 shows U.S. automobile factory yearly sales from 1900 to 1970 are taken from the U.S. Bureau of the Census (1975, Series Q148, p.716). The years 1950 to 1981 of the top line (domestics) and the entire lower line which is automobiles imported into the U.S. per year from 1950 to 1981 are taken from the U.S. Bureau of the Census (1982, No. 1060, p.615).



Note that for 1950 to 1970 the imports were accelerating by a factor of x3.5 (250% increase) every five years! At that acceleration, imports would clearly have half the U.S. market by 1990. The U.S. Manufacturers did not make such charts. They merely monitored their own sales, which, in spite of yearly rhythms, were actually accelerating at a factor of x1.1 (10% increase) every five years. Not so bad, if that is all you look at. But, terrible if you realized that the competition is accelerating over three times faster!

This example of back-charting an acceleration target (U.S. made cars) with its deceleration target (foreign made cars) shows the overall picture of automobile manufacturing on the U.S. society. It is clear that in the global marketplace it is no longer enough to monitor your own corporate sales or even your own and your national competitors sales. It is now necessary to continually monitor your international competitors' sales as well to get their macro level picture.

This major tragedy in the U.S. economy could have been avoided if managers had merely monitored macro level impact on appropriate multiply scales, charting their own sales against their foreign competitors.

Example of Mega Level Monitoring - Institution Readmissions

From the 1960's through the 1970's institutions for the developmentally disabled (called mentally retarded at that time) had a marked acceleration in residents discharged (live releases) and a marked increase in expenditures. However, their effect on society at large was worsening because their readmissions were accelerating more rapidly than their releases. The residents they released as ready for community living were not ready because more and more were being taken back to the institution each year.

Figure 5 displays the maintenance expenditures in dollars, and the total number of residents, personnel, admissions, and discharges (live resident releases) for the total number of U.S. public institutions for the mentally retarded from 1936 through 1975 (U.S. Bureau of the Census, 1975, Series B 428-443, p.85). These frequencies are charted on a yearly standard celeration chart. Also charted are the readmissions of former residents for 1951 through 1974 (Conroy, 1977).

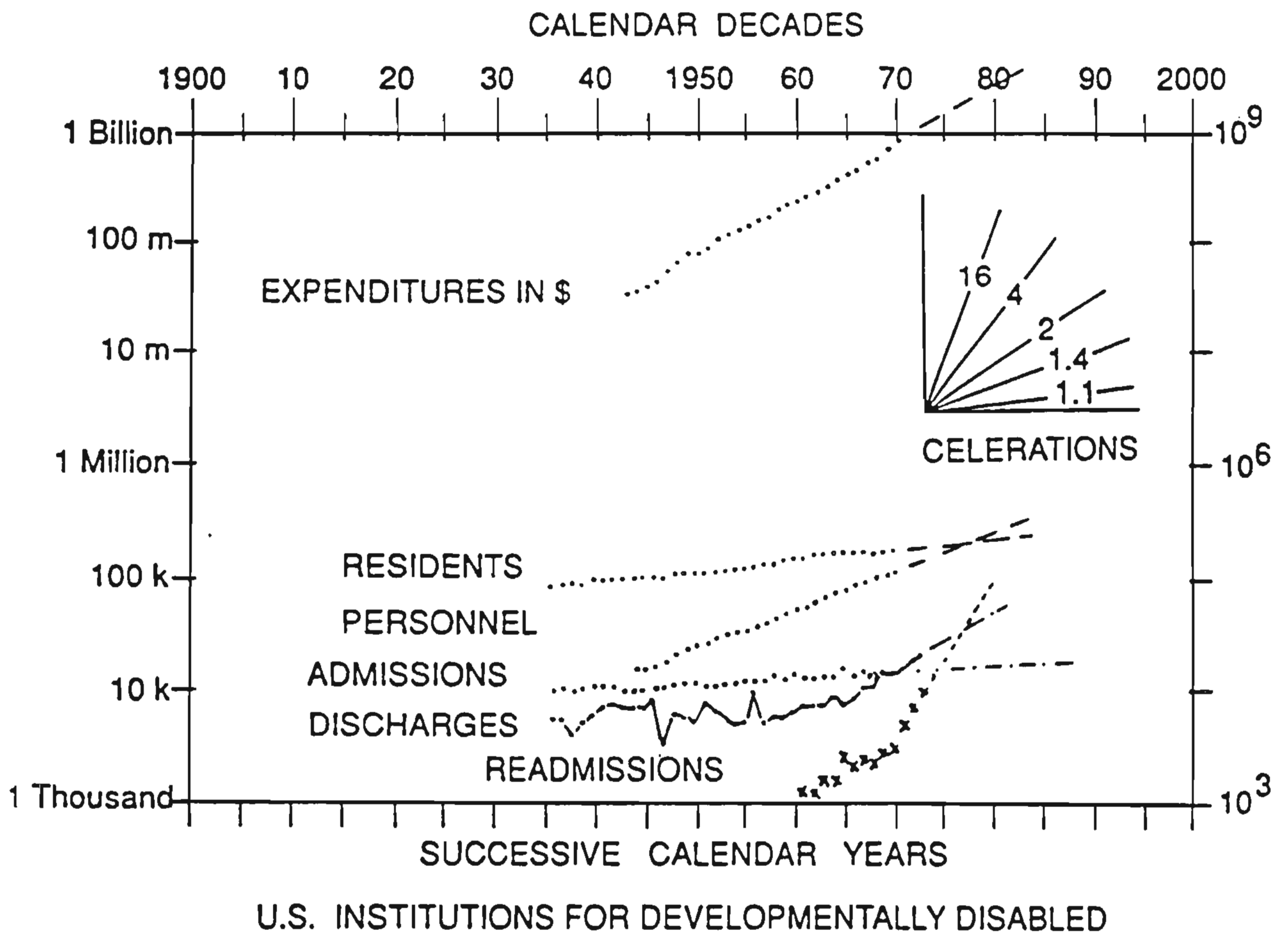


Figure 5. Discharges compared with readmissions for U.S. institutions for the developmentally disabled per year on a standard celeration chart. Number of admissions, residents, personnel and total expenditures per year are also charted to make interpretation easier.

As you look at this chart, squint your eyes and look at six lines at once. Notice that the top three lines are almost straight. There is almost no bounce in these frequencies, indicating that the institutions size (cost, residents, and staff) is regularly multiplying. This means these values can be projected by straight lines on the chart with high accuracy. Notice that admissions are almost as straight with higher admissions in 1958, 1960, and 1965.

Note that residents and admissions are parallel accelerating at x1.1 (+10%) every 5 years which is the same growth for the U.S. population at large. The lines are about x11 apart, indicating that about one out of 11 of the residents was admitted during the prior year. If we had charted U.S. Population we would have seen that about 1 out of 1,000 U.S. persons is in an institution for the developmentally disabled, because the population line would have been x1,000 above the residents line.

Note that the expenditures line is much steeper (x1.8) than the residents line indicating that the cost of personnel is probably a major factor in expenditures, since the number of residents isn't parallel with expenditures. The distance of the expenditures line above the residents line has increased from 1947 to 1970, indicating higher cost per resident. The distance increased from x300 or $300 per resident in 1947 to x5,000 or $5,000 per resident in 1970.

Note that the personnel line (x1.5) is steeper than the residents line (x1.1). This means that the number of staff per patient is multiplying at 1.5 divided by 1.1 equals x1.4 every 5 years (+40%). The ratio went from 1 staff for 7 patients in 1945 to 1 staff for 2 patients in 1965. Is this due to inexorable, creeping bureaucracy, or actually an improvement in the quality of treatment? An actual improvement in treatment would produce more discharges. By looking at the discharges we can answer this question.

The next to the bottom connected line is discharges (called live releases in the U.S. handbook) Note that the line is more bouncy than the other lines. This indicates the production of the institutions is not regular, and could be subject to political pressures other than financial. They could not be due to financial causes or the expenditures line would also bounce proportionally. The discharges have bounced as high as x2.5 from lowest to highest years. There is an unexplained lemony year in 1947 and an unexplained exemplary year in 1956. These should have been examined for special causes, but obviously were not.

From 1939 to 1965 the discharges multiplied by x1.1 every 5 years, parallel with admissions, residents and the general population growth. This represents stable institutional production. During this time personnel were multiplying at x1.3, but they had no effect upon discharges. This favors the creeping bureaucracy rather than increased quality of treatment interpretation. Discharge acceleration shifted abruptly in 1965 from x1.1 to x1.3 which was maintained through 1970. Since there was no change in personnel at this time, the shift is probably due to the de-institutionalization, community-treatment push that started in 1965. How effective was the increased acceleration in discharges which now seems to parallel personnel?

The quality of the institutions' production is revealed by comparing discharges with readmissions. The bottom line of small "Xs" is readmissions. The readmissions are not regularly reported to the government, and were obtained from a different source (Conroy, 1977). If readmissions also go up with discharges, then the treatment is not of higher quality. For then the institutions are merely pushing more residents out their doors who cannot survive in society at large, and who later must be readmitted.

Note that from 1960 to 1969 readmissions accelerated at x1.3 parallel with discharges, showing that the quality of discharge was maintained. The readmission line was about x5 below the discharge line showing that for every five discharges there was one readmission. Then in 1969 the readmissions abruptly turned up to x6 every five years. There was no concomitant turn or jump in discharges or personnel. By 1972 for every 2 discharged there was one readmitted. This rapid decrease in the quality of discharge is unexplained.

If you have followed these 70 lines of text and Figure 5, you have understood the 500 numbers in 16 columns and 35 rows of the table on page 85 of the U.S. statistical handbook.

Any way you look at it, the production track record of the institutions is very poor. But how could it be any different, since we pay them for number of residents warehoused? It would be very different if institutions were paid for number of residents permanently discharged. Here again, it looks like society gets what it pays for!

Quality Navigation™

Precision teachers and learners have self-monitored academic performance on standard celeration charts in classrooms and learning centers since 1965. Over 32,000 learning charts collected in a main-frame computer demonstrated that human performance frequencies increase by multiplying and decrease by dividing. A large number of laws and rules of learning were discovered over these 25 years of standard charting of academic performance.

We have long known that business performance frequencies (attributes) followed the same rules. In 1991 I found that business product continuous values (e.g. ohms resistance) also follow the same rules and can be charted easily on the same standard celeration chart. The chart we had developed for pre-schoolers could be used by blue collar workers to monitor both values and attributes easier and faster than Statistical Process Control with much less training. I am offering these methods to industry for quality control under the trademark "Quality Navigation™."

JOB AID FOR MONITORING PERFORMANCE

I close this chapter with the following procedure table to serve as your job aid for monitoring performance frequencies:

Step	Action
1	List your products that help your customers do **their** job.
2	Reword list to noun-past tense verb form.
3	Discard all behaviors that do not pass "dead-man test."
4	Discard all accomplishments that do not pass "leave-it test."
5	Pinpoint acceleration/deceleration counting pairs.
6	Count each in real time (minutes, days, weeks, months, or years).
7	Chart original frequencies in real time on standard multiply charts.
8	Monitor for exemplary and lemony outliers.
9	Analyze work place for outlier causes.
10	Institute exemplary causes. Eliminate lemony causes.
11	Monitor for early signs of trend changes.

Use it well.

REFERENCES

All, P. (1977). *From get truckin' to jaws, students improve their learning picture.* Unpublished Master's Thesis, University of Kansas, Lawrence.

Bijou, S. W. & Baer, D. M. (1961). *Child development: Vol 1. A systematic and empirical theory.* New York: Appleton-Century-Crofts.

Binder, C. V. (1990). Closing the confidence gap. *Training*, September, 49-56.

Binder, C. V. & Bloom, C. (1989). Fluent product knowledge: Application in the financial services industry. *Performance and Instruction*, February, 17-21.

Conroy, J. W. (1977). *Mental Retardation.* August 1977, 44-46.

Cordiner, R. J., (1956). *New frontiers for professional managers.* New York: McGraw-Hill

Crosby, P. A. (1984). *Quality without tears: The art of hassle-free management.* New York: McGraw-Hill.

Daniels, A. C. (1989). *Performance Management.* Tucker, GA: Performance Management Publications. 250 Pages.

Deming, W. E. (1986). *Out of the crisis.* Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Duncan, A. D. (1971). The view from the inner eye: personal management of inner and outer behaviors. *Teaching Exceptional Children, 3* (3), 152-156.

Evans, R. I. (1968). *B. F. Skinner: the man and his ideas.* New York: E. P. Dutton & Co., Inc.

Geis, G. L., & Smith, M. E. (1992). Chapter 9: The function of evaluation. In H. D. Stolovitch & E. J. Keeps (Eds). *Handbook of human performance technology.* (130-150). San Francisco: Jossey-Bass.

Gere, J. M. & Shah, H. C. (1984). *Terra non firma.* Stanford, CA: Stanford Alumni Association.

Gilbert, T. F . (1978). *Human competence: engineering worthy performance.* New York: McGraw-Hill Book Co.

Gilbert, T. F. (1992). Foreword. In H. D.Stolovitch & E. J. Keeps (Eds), *Handbook of human performance technology.* (pp.xiii-xviii). San Francisco: Jossey-Bass.

Gitlow, H. S., & Gitlow, S. J. (1987). *The Deming guide to quality and competitive position.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Gove, P. B. (Ed.) (1961). *Webster's third new international dictionary of the english language (unabridged).* Springfield, MA: G. & C. Merriam Co.

Haughton, E. C. (1974). Define your act and set your fluency goals in personal, social, and academic areas. *Special Education in Canada, 48* (2), 10-11.

Haughton, E. C. (1981). REAPS. *Data Sharing Newsletter*, March, Waltham , MA: Behavior Prosthesis Laboratory.

Holland, J. G. & Skinner, B. F. (1961). *The analysis of behavior.* New York: McGraw-Hill.

Holzschuh, R. D. & Dobbs, D. (1966). *Rate correct vs. percentage correct.* Unpublished manuscript, Educational Research, University of Kansas Medical Center.

Johnson, K. R. (1989). *Executive Summary.* Morningside Corporation. (810 18th Ave., Seattle, WA 98122).

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist. 47*(11), 1475-1490.

Kaufman, R. (1992). *Strategic Planning Plus.* Newbury Park, California: Sage Publications.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 1-55.

Likert, R. (1958). Measuring organizational performance. *Harvard Business Review, 36*(2), 41-50.

Lindsley, O. R. (1956). Operant conditioning methods applied to research in chronic schizophrenia. *Psychiatric Research Reports, 5*, pp.118-139.

Lindsley, O. R. (1991). From technical jargon to plain English for application. *Journal of Applied Behavior Analysis, 24*, 449-458.

Lindsley, O. R. (1992a). Why aren't effective teaching tools widely adopted? *Journal of Applied Behavior Analysis, 25*, 21-26.

Lindsley, O. R. (1992b). Precision Teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*, 51-57.

Lindsley, O. R. (1992c). *Skinner on Measurement:* Kansas City, KS: Precision Media.

Mager, R. F. (1992). Afterword. In H. D. Stolovitch & E. J. Keeps (Eds), *Handbook of human performance technology.* (pp.743-746). San Francisco: Jossey-Bass.

Meadows, D. H., Meadows, D. L., Randers, J ., & Behrens, W. W. (1972). *The limits to growth.* New York: Universe Books.

Pennypacker, H. S., Koenig, C. H., & Lindsley, O. R. (1972). *Handbook of the standard behavior chart.* Kansas City , KS: Precision Media.

Shewhart. W. A. (1939). *Statistical method from the viewpoint of quality control.* Washington, DC: Graduate School, Department of Agriculture; reprint edition, 1986, New York: Dover.
Skinner, B. F. (1938). *The behavior of organisms.* Englewood Cliffs, NJ: Prentice-Hall.
Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis.* New York: Appleton-Century-Crofts.
Skinner, B. F. (1983). *A matter of consequences* New York: Alfred A. Knopf.
Smith, M. E. & Geis, G. L. (1992). Chapter 10: Planning an evaluation study. In H. D. Stolovitch & E. J. Keeps (Eds). *Handbook of human performance technology.* (151-166). San Francisco: Jossey-Bass.
Stolovitch, H. D. & Keeps, E. J. (1992). (Eds). *Handbook of human performance technology.* San Francisco: Jossey-Bass.
Thurstone, L. L. & Chave, E. J. (1929) *The measurement of attitude.* Chicago, IL: University of Chicago Press.
U. S. Bureau of the Census. (1975). *Historical statistics of the United States, Colonial times to 1970, Bicentenial edition, Part 1.* Wahington, DC.
U. S. Bureau of the Census. (1982). *Statistical Abstract of the United States: 1982-83.* (103rd edition.) Wahington, DC.