To Carl
With Great Respect
Much Love Og -09

FROM TRAINING EVALUATION TO PERFORMANCE TRACKING[1]

OGDEN R. LINDSLEY[2]

1649

CONTENTS PAGE

[1] Accepted for publication as Chapter 11 in Stolovitch, H. D. & Keeps. E. J. (Eds.). (1998). The Handbook of Human Perfomance Technology, Edition 2. San Francisco: Jossey-Bass.

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

FROM TRAINING EVALUATION TO PERFORMANCE TRACKING

As Human Performance Technology (HPT) shifts from training to performance improvement, measurement is forced to shift from training evaluation to performance tracking. For forty years, since 1959, the four levels of I reaction, II learning, III behavior, and IV results (Kirkpatrick, 1994) have ruled our thinking and practice in training evaluation . Over the same period, out of 100 programs 100 have measured reaction, 70 learning, 50 job behavior, and only 10 measured business results (Phillips, 1994, p. 8). *Training* magazine's 1996 survey of 1400 companies reported that 86 out of 100 courses measured reactions, 51 measured learning, 50 measured behavior, and 44 measured results (Industry report, 1996, pp. 36-79). This reports a big step forward in measuring the long neglected business results.

RECENT MOVES TO TRACKING IMPACT

Now, at last, we are moving not only to business results but beyond - to tracking organizational impact. At many places we see signs of this shift in the focus of evaluation. Performance tracking is on the move. Performance tracking promises to be fun and exciting and useful for a change. The following four recent examples will whet your appetite.

MEASURIT's *Tracking Training Impact.* In MEASURIT's (1996) two-day seminar program, *Tracking Training Impact,* the Five Level Tracking Model defines assessment levels used to quantify training impact. In Level One critical business issues and needs are assessed before investing in training. This identifies the most cost-effective solution to the presenting problem. All following tracking levels, including Level Five, which assess training's worth to the organization, are linked to resolving the business needs identified in Level One.

Esque and Patterson's *Getting Results.* For years we have needed collections of case studies to give us ideas on how we might track performance improvement in our projects. Of the 22 performance improvement case studies reported in this book, seven (over one third) reported results that impacted the organizational level. Two of these reported impact in dollars of revenue. Twelve cases reported improved job performance. Three reported learning from the training, and none reported learner reactions to training (Esque & Patterson, 1998).

Brethower and Smalley's *Performance-based Instruction.* In their recently published book, the authors write, "Evaluation, from front-end through impact, is part of performance based instruction. In fact, it is the only form of instruction in which Level IV evaluation is routine," (Brethower and Smalley, 1998).

Spitzer's *Super-Evaluation.* Author of the best selling management book, *SuperMotivation* and an IBM Corporation consultant, Spitzer is now presenting a novel approach to evaluation he calls "*Super-Evaluation*" at workshops and in consulting engagements (Spitzer, 1998). He argues that evaluation is usually done entirely "after the fact." Spitzer suggests that the very first thing we should do when starting a project is to select the desired impact of the project on the organization and use evaluation to mold the project to that desired outcome, rather than using evaluation to simply assess the consequences at the end. The desired impact decision should always come first and the project should be designed to achieve that outcome. Throughout the performance improvement project, the impact goal guides the project and is continuously monitored.

MEASUREMENT, MONITORING, AND TRACKING

To clarify this shift from conventional measurement to performance tracking measurement, monitoring, and tracking must be clearly distinguished. They must not be considered merely new words chosen to escape negative reactions to the word evaluation. These three terms must be clearly defined to make our evaluation alternatives clear. All three are ways to put numbers on performance. These numbers let us see whether our project procedures are improving performance less and less or more and more. We can also use these numbers to see which one of our methods had the biggest and quickest effects with which one of our clients.

Measurement

Measurement is done before, occasionally during, and usually after the performance improvement project. It includes a set of actions to take and is always outside the system being measured. Measurement is what is taught in most conventional measurement and evaluation courses and workshops. Measurement is strong on the theoretical and statistical and weak on the practical. A measurement system is added to the performance being measured. Measurement is not continuous in real, calendar time.

Since measurement is not continuous, but usually is done only a few times during a project, workers never adjust to it. Workers react to measurement negatively, as if it were a test. In their experience, tests have most often been followed by punishment rather than reward.

Monitoring

Several years ago the writer described the many differences between measuring and monitoring, but did not separate monitoring into external monitoring and self monitoring or tracking (Lindsley, 1997a). Most of the examples listed under monitoring would now be listed as tracking examples.

Monitoring systems are always outside the performance being improved. Monitoring differs from measurement by being continuous. It records all the time that performance occurs. Monitoring requires an external recording system that is designed and added to the performance change system. Monitoring systems collect performance numbers as they happen, so that corrections can be made in real time before the performance gets too far off course.

Because monitoring systems are clearly outside of the performing system, they are often suspect and reacted to negatively. The word monitor itself has negative connotations from school days (hall monitor, playground monitor).

Tracking

Tracking occurs when a counter within the performing system records itself without interfering with its performance. Ideally, the performance itself leaves a track. Like rabbit tracks in the snow, the performance records itself. Tracking gives the most accurate and most sensitive performance numbers. Tracking costs the least and disturbs the performance the least.

Tracking has face validity and is accepted by all workers and managers as an important indicator of performance improvement.

Monitoring and Tracking Make Baselines Unnecessary. Because both monitoring and tracking continuously record performance, trends are clearly displayed. Clear trend displays make baselines unnecessary because jumps and turns in the trend line produced by your performance improvement methods are clearly seen. Baselines still add information about the trends before the performance improvement project, but are no longer required to demonstrate method results.

KEEP IT SIMPLE

A related shift from academic and statistically sophisticated measurement to simple practical methods easy for workers to use on the shop floor has occurred. This shift away from the barren, hard-to-learn academic measures occurs even in first level participant reaction evaluations. Here is a recent example.

ISPI From Likert Scale to Check List. For decades The International Society for Performance Improvement (As NSPI, and ISPI) used a traditional standard nine item Session Evaluation with a five point Likert scale. Participants circled a point for each item to record their reactions to the annual conference presentations. It was simple, it was standard, and it got done. The circled numbers on the ordinal Likert scales were averaged. (This bothered some of us purists because it was mathematically incorrect - medians would have been more accurate for the ordinal data). The averaged results were interpreted by ISPI's program committee to make decisions about inviting presenters to encore or to qualify to present a different topic at the next year's conference.

For the 1998 convention ISPI changed its Session Evaluation form to a 21-item check list. The items on the check list describe program committee interests directly. Simple totals of the number of checks to each of the 21 items give much more accurate information to the program committee than did the Likert scale averages of nine questions.

Using the old Likert scales, the ISPI program committee had to interpret when the overall average of the nine items was 4.67 to determine whether to invite the presenter to encore at the next convention. Now, when the committee reads that 62 out of 68 participants checked the 20th check list item, "I believe this session should be presented next year in the Encore track that repeats the best sessions at this conference," interpretation is unnecessary. Simpler is better.

KISSING PERFORMANCE TRACKING HELLO

This shift to the new performance tracking involves more than just simplification. In aviation cadet training during World War II, I first heard KISS meant "Keep It Simple, Stupid!" But for performance tracking we must add four other features to "Simple." KISSING helps us remember the five features required to make performance tracking powerful: KISSING means Keep It Simple, Standard, Impactful, Natural, and Graphic.

Keep It

S for Simple

Our performance tracking system should be simple enough, so that if it has to be taught, it only has to be taught once.

In a recent interview, Norman, Head of the Appliance Design Center in the Consumer Products Group of Hewlett-Packard, said, "My goal is not to have zero training, but to have one-time training. If I don't quite understand a computer system immediately ... and someone shows me, ... I never have to have it explained again." (Dickelman, 1998, p.37).

Most of us have had to have conventional evaluation methods explained to us over and over and over again. By our new simplicity requirement this is unacceptable.

Job aids and check lists are simpler than flow charts and rating scales. Recent books on performance technology (HPT) are full of check lists and case examples. Fuller's (1997) *Managing Performance Improvement Projects* has many realistic case studies, and many assessment tools and checklists to simplify things for readers. Brethower and Smalley (1998) include in their *Performance-based Instruction* many examples, job aids, and check lists, and even include a computer disk to make printing the check lists easy by eliminating reader keyboarding or copying.

Most of us spent hard years and hard cash in universities learning to write and talk in four and five syllable academic jargon. Academic talk is the hallmark of the learned professional specialist. With so much invested in our jargon it is very hard for us to talk and write plain English to our clients and their workers. Behavior analysis terms had to be translated into plain before Precision Teaching could be widespread (Lindsley. 1991). Mager (1986, p. 97) gives his manuscripts to two twelve-year-olds and asks them to draw a circle around any word they do not yet know. He reports that a third of the circled words can be replaced by words that everyone knows. Mager's books are so beautifully clear that they are in great demand and have been easily translated into many foreign languages.

Our information explosion has put even greater pressure on using short, clear, plain English words. Evaluation has five syllables: Tracking has two. "Sudden shift in level" has six syllables: Jump has one. "Change in trend" has three syllables: Turn has one syllable. Authors are even dropping their middle initials (D. M. Brethower and K. A. Smalley to Dale Brethower and Karolyn Smalley). Other authors are publishing with their one syllable nick names. Al Reis and Jack Trout (1993) authored the best selling *The 22 Immutable Laws of Marketing*. I am sure that mother Reis named her son Albert and he probably also had a middle name. Aim for one syllable words.

In his presentations on Super Evaluation, Spitzer (1998) quotes Albert Einstein as follows, "Everything should be as simple as possible ... but no simpler."

S for Standard

Every minute of every day our lives are governed and made easier by standards. Our buildings are made of beams of standard width and thickness. Our hats, dresses, trousers, shoes, and rings come in standard sizes. Everything we use - hand tools, furniture, utilities, appliances, vehicles, and computers are designed, built, and used by standards. Standards are vital to our society.

We use standard terms, methods, and analysis procedures in Human HPT. We use the standard seven, plus or minus two, in making our training lists and job aids (Miller, 1956). However, we do not use standard evaluation measures. Much has been written on the need for educational standards, but little real standardization has been done (Dean, 1994).

Imai (1997, p. xviii) tells how the Japanese *Gemba Kaizen* (workplace guality and production control) has three ground rules: 1. Housekeeping, 2. Waste elimination and 3. Standardization. Good housekeeping cuts the failure rate in half. Standardization cuts that failure rate in half again to one quarter of the original failure rate.

The 5 point Likert scale came close to being a standard, and that is why it got done. But it produced very poor numbers that did little to improve performance. Percent is also very popular and close to being a standard. But percent is also a very crude and weak performance measure. Percent only describes the relationship between two other things. It ignores the sizes of those two things. The dangers of percent are described in the following section labelled natural.

HINIBUs and Egos Block Standard Measures. HINIBU stands for Horrible If Not Invented by Us. It is a disease found in universities, small business, and even some large corporations. People infected with HINIBU cannot try anything new unless they have changed it enough to make it look like they invented it in house. We all know that we should put our client's corporate logo on the reports and materials we develop for them. However, we must be very careful not to let corporate personalization change the tracking system to the point that it is no longer standard.

Frequency is Universal. Without a doubt the only possible across-the-board performance standard is frequency - how many happened in how much time. Each happening of everything in the world that happens can be counted. All counts cover a counting time. Every count divided by its counting time gives a frequency. Therefore one thing's frequency can be used to compare it with other things like it, or things very different from it.

A few examples follow. The average six year old laughs 300 times a day. Adults laugh between 15 and 100 times a day. Therefore children laugh over three times more than adults. The normal blink rate for someone speaking on TV is 31 to 50 blinks per minute (bpm). In the presidential TV debates, Clinton averaged 99 bpm, and Dole averaged 147 bpm. So, from observing blinking, Clinton appeared twice as nervous, and Dole three times more nervous than most TV speakers.

Almost always track quality by counting how many good ones and how many bad ones happened, and separately charting them.

We should use the natural counting times of our daily life and work. These times are number per minute, per hour, per day, per week, per month, per quarter, and per year.

B. F. Skinner (1950) gave us frequency saying, "Rate (frequency) is a universal datum." By this Skinner meant that everything in the universe has a frequency that can be counted and occurs in time. Therefore frequency provides a comparison standard.

Frequency is a Dimension of Performance. Research has taken us beyond Skinner in demonstrating that frequency is a dimension of performance. This means that changing the frequency changes the performance. Try this little experiment to convince yourself of this fact. Take a plain piece of paper and write down your starting time. Now write your signature as slowly as possible. Just barely move your pencil, writing letters as slow as one or two a minute. Keep slowly writing your signature. After 10 minutes stop and look at your signature. your second or third grade signature should be wasritten. It was in you all these years, stored at a very slow frequency. Now write your current signature as many times as possible in one minute. Count up the number of letters written per minute in both conditions. Your third grade signature should be at 2 to 4 letters per minute. Your current signature should be at 150 to 220 letters per minute.

Change the frequency of writing and change the signature form. Change the frequency of light waves and change the color of the light. Change the frequency of sound waves and change the tone of the sound. Just as frequency is a dimension of light and sound, frequency is a dimension of performance. To fully describe a performance, its frequency must be described.

This is why training people to high frequencies of performance makes them fluent. The form and control of the performance changes at high frequencies. Fluency produces more retention, more application, more stability, and more confidence (Binder, 1990, 1996). The guidelines and demands that fluency places on instructional design have been recently detailed (Lindsley, 1997b).

In short, when it is standard it is easier to do right. When it is standard it gets done. The best standard for performance is frequency.

I for Impactful

The first six paragraphs of this chapter describe the shift in the focus in training evaluation from Kirkpatrick's (1994) four levels to organizational impact. I described four recent examples: MEASURIT (1996), Esque & Patterson (1998), Brethower & Smalley (1998), and Spitzer (1998) to demonstrate this shift to impact.

When we shift to impact we must be careful to choose frequencies that we can use as guides to improve our project while we go along. The majority (47 out of 58) of the projects included in three collections of impact projects (Phillips, 1994, 1997; Esque & Patterson 1998) report only before and after impact measures. These justify the project to management, but cannot guide performance improvers during their project.

Only 11 of these 58 projects (1 out of 5) included continuous tracking data. Of these 11, 8 tracked months, 1 tracked quarters, and 2 tracked years. These times are not short enough to accurately guide project improvement. Weekly frequencies are better and daily frequencies are best for continuous feedback to guide workers. A daily chart appeared in Esque and Patterson but it charted "cumulative percent of quota" which made it impossible to reclaim the original frequencies for comparing or re-charting.

Phillips (1994, 1997) advocates using Return On Investment (ROI) to measure organizational impact. The ROI % equals (net program benefits divided by program costs) times 100. The suggested advantage is that chief executives will have training impact reported to them using the same financial figures that they used for their other investments. Because most ROI cannot be computed until after the project it cannot guide the project en route. Therefore, it is an impact measure that seldom can be used for tracking.

In summary, only one fifth of our published projects that measure impact track the impact continuously enough to guide project improvement. Of those that do, almost none track the weekly and daily frequencies that most effectively improve projects.

N for Natural

Original natural numbers that your performance system kicks out should not be "cooked" or transformed. This is the major strategic error made by most management information systems designers. They can not leave nature alone. They cook the original numbers in attempts to focus your attention on relationships between the original numbers. Cooking vegetables loses much of their original flavor and texture. Cooking original performance frequencies loses the details and sensitivity to change needed to guide performance improvement projects. Examples of cooked data are: Percent rejected, percent or proportion of standard, or product quality index (Esque & Patterson, 1998, p. 44). A complete list of data cookings would be very long.

The Dangers of Percent. Of the twenty one data charts in Daniels' (1989) *Performance Management* nine are percents. Of the seven performance data series in Daniels' (1994) *Bringing out the best in people* five are percents. Nine of the twenty two cases reported in Esque and Patterson's (1998) *Getting Results* measured percents.

Percent is often your client's favorite performance measure. However, percent is insensitive to changes in performance and actually dangerous to use. Skinner was aware of the problems with percent when he wrote "Do not spend time on articles in which graphs show changes in the time, or number of errors to reach a criterion, or percent of correct choices made" (1969, p. 93). Holzschuh (1966) spent two years of full time postdoctoral research comparing the sensitivity of percent correct with frequency correct to classroom curriculum changes. He concluded "Percent is the worst thing that ever happened to education." One of my most successful workshops is titled *The Dangers of Percent and How to Avoid Them* (Lindsley 1994).

An impact measure that your client has long recorded often will be a percent or a ratio of some standard or a percent of a company aim or quota. When given a percent try to locate the original numbers from which the percents were calculated. The original, natural numbers are best to track performance improvement throughout the project. Because clients often love their percents, do not ask clients to throw their percents away. Just locate the original numbers for tracking performance. Then share both the originals and the percents with your client.

People think they understand and know percents, but they do not. Eight out of ten of the errors in the mathematics section of standard achievement tests made by both children and adults are in calculating and interpreting percents Less than half of pre-service teachers scored higher than 50% correct on a test of percent problems (Parker & Leinhardt, 1995). Percent is hard to use because it uses the add language of more than, less than, increased by, and decreased by, which both hides the multiply meanings of percent and suggests a symmetry that is not really there. Because percent is so hard to use, errors in calculating percents appear in many professional publications. There is a percent calculation error in the data reported in Esque and Patterson's *Getting Results* (1998). There is another percent calculation error in Phillips' *In Action Vol. 1* (1994). Try to find them.

Averages Destroy Order. Shewhart, who originated Statistical Process Control (SPC) at Bell laboratories, wrote, "It is well to keep in mind that numbers and order are the two aspects of original data that are amenable to mathematical analysis" (1939. p. 90). If the traditional root-mean-square formula for the standard deviation is used to determine the upper and lower control limits, the control charts are not sensitive to variations. This happens because the order in the original data series is lost by the averaging process. The formula used to set SPC control limits retains the order from the original data and is sensitive to variance outliers. Shewhart (1939, p. 90) went on to say, "It was the order that furnished the clue to the presence of assignable causes of variability that were later found and removed."

This means do not average original data. Keep original frequencies intact. According to Shewhart, averaging destroys order, which is half the value in your data. This also means that graphs are necessary to maintain the order in original data series. This also means that time series graphs display changes more sensitively than tables and other graphs.

G for Graphic

A picture is worth a thousand words. Research has shown: Charts have higher impact on readers than text based pages, recall and comprehension are higher with charts, and complex data relationships are more easily shown with charts. Daniels (1994, p. 101) writes "Employees on production jobs, office jobs, and even creative jobs in software development can easily arrange a system of daily feedback. Once again I am referring to graphic feedback."

Imai (1997, p. 249) reports the power of company wide visual management. He states "Visual management means displaying ... various graphs and charts on the current status of Kaizen activity on the walls and in every corner of the factory."

Imai (1997, p. 114) also describes *asaichi* (morning market) in which the first thing every morning before work the rejects from the day before are displayed on a table and countermeasures adopted on the spot. This fits Tosti's (1978) and Lindsley's (1995) ideas of telling workers what to do and what not to do just before their next chance to perform. Therefore charts displaying yesterday's performance should be shared and discussed by the worker group the first thing at the start of each work day.

Unfortunately graphical power is not yet used by most performance technologists. They visually display many flowcharts and diagrams of their procedures, but almost no data time series charts of their effects. The *Handbook of Human Performance Technology. Vol. 1* (Stolovitch & Keeps, 1992) contained 817 pages, 44 chapters, but only 2 data time series charts. If the 6 charts in this chapter do not get edited out, it alone displays 3 times more data than the whole of the earlier handbook.

FILL THE FRAME CHARTS

Just as it is easy to lie with statistics (Huff, 1954), it is easy to lie with charts unless the charts themselves are standard. Most of us were taught to make charts by stretching our data to fill the frame of a chart. We made a rectangle with number up the left side and time across the bottom. No standards were given us for what size the rectangle should be, what numbers should be used up the left, or what time (minutes, days, weeks, months, or years) should be across the bottom. Most of us naturally draw a rectangle just big enough to contain our data points. Most of us also made the numbers add up the left and across the bottom.

These fill the frame charts are what almost all computer graphing programs make for us. The user has no control over the exact size of the chart frame, so the horizontal and vertical frame proportions cannot be set. The range of the data values controls the size of the horizontal and vertical scales of the frame.

These fill the frame charts maximize seeing your data details, but make small changes in one chart look just as big as large changes in another chart. Stretching content to fill the frame makes a mouse look like just as big and more formidable than an elephant

A fill the frame chart with an add scale up the left and a fill the frame chart with a multiply scale (log to base 10) up the left for each of six different data sets appear below. Look at them to see how the data are stretched to fill the frames.

STANDARD CHANGE CHARTS™

Ideally, performance improvement in industry should be tracked on standard charts just as our students in Precision Teaching classrooms have done for thirty years. Some of the facts that set the stage, the features, and the benefits of these standard charts follow.

Facts That Set the Stage. Five facts from three different sources set the stage for designing standard charts to track performance change.

1. Frequency can be used to track any and all performances.
(Skinner, 1950).
2. Every performance changes by multiplying or dividing.
(Meadows. et al., 1972).
3. Ratio charts best show rate of change and percentages.
(Schmid, 1954, p. 109).
4. Standard chart slopes make reading change easy.
(Skinner 1938).
5. Self charting makes performance tracking affordable.
(Skinner 1938).

These five performance and charting facts were combined in designing a standard slope chart in 1965.

Although committed to education, but still not entirely free from an experimental psychology background the standard chart was first named for what we charted on it - the Standard Behavior Chart (Pennypacker, Koenig, & Lindsley, 1972). Later, it becane clear that it was the slope of the chart that was standard, not the use of it. Also, users started charting many other things than behavior. Then the writer changed the name to Standard Celeration Chart which is still its technical term. Frequency is up the left of the chart and the standard slope is change in frequency or celeration. Upward slopes show acceleration. Downward slopes show deceleration. Celeration had to be coined.

Standard Change Chart™ and the Change Factor Fans™ are copyrighted and are trademarks of Behavior Research Company for commercial use. Paper Standard Change Charts, overhead transparencies, computer templates, and on site workshops are available from Behavior Research Company, PO Box 3351, Kansas City, KS 66103, Fax 913-362-5900. Price lists are available at www.onlearn.com/brco.html.

Standard Change Chart Dimensions. To fit on slides, overhead screens, and standard 8 1/2 by 11 inch notebook paper, the full size paper chart frame was made 8 inches wide and 5.3 inches high. To put most of a school semester on one sheet the horizontal time add scale covered 140 days or 20 weeks. To include the full range of human daily performance frequency a vertical, times 10 multiply scale (base 10 log) spread from one per day (.001 per minute) to 1,000,000 per day (1000 per minute).

Standard Change Chart Slopes. The chart dimension proportions also were carefully selected so that a line drawn from the lower left corner to the upper right corner indicates a doubling (times 2 or x2) in performance each week. This is an angle of 34 degrees. A line drawn from upper left to lower right corner of the frame shows a halving (divide by 2 or /2) of performance each week. The angles and meanings are symmetrical. Performance changing half as steep, going only half way up the chart in 20 weeks shows times 1.4 growth per week (an angle of 19 degrees). Performance changing twice as steep as times 2 would go all the way to the top of the chart in only 10 weeks, multiplying by times 4 each week (an angle of 53 degrees). Performance changing even steeper would go all the way up the chart in only 5 weeks, multiplying by times 16 each week (an angle of 69 degrees).

If the proportions of these dimensions are kept standard, the SCC can be enlarged or reproduced to any size for presentation and publication without changing the angle and meaning of the slopes and change factors, What is standard is the slopes and their meanings, not the physical size of the chart frames.

These chart frame dimensions and change factor angles are maintained for charts covering the different levels of time. The daily chart, weekly chart, monthly chart, and yearly chart all have the same proportions and same change factor angles. This means that learning one set of performance change factors works with any and all levels of time and corporate organization. The worker who has learned her daily chart easily reads and understands her supervisor's weekly chart, her manager's monthly chart, and the yearly financials in the company annual report.

The performance decay factors are the same except they go down the chart rather than up. This symmetry makes it very easy to learn both growth and decay factors at once. These bench mark change factors are easily learned by adults in a one-day standard charting workshop. Having learners stand in the room and hold their arms out at the correct angle while repeating the factor as a leader calls out change factors at a pace of 30 per minute, helps them rapidly feel and learn these chart slopes. First grade school children learn them in a few weeks of charting ten minutes a day.

Standard Self-Improvement Charting. Skinner (1938) taught his rats to produce their own standard performance frequencies on his cumulative response recorders. The standard slopes of these records displayed performance rate or frequency. Following this lead, precision teachers taught school children to chart their own daily performance on Standard Change Charts™ (Lindsley, 1971). The standard slopes of these charts displayed performance change or weekly learning. Plotting hits and misses on their standard chart let each child not only track their own daily performance and accuracy, but also track their own weekly hit and miss learning (Pennypacker, Koenig, & Lindsley, 1972). Hit learning is independent from miss learning. When their learning slopes were shallow, the child changed something, or asked help from another student. If nothing helped, the child called on the teacher for learning advice.

Standard Effectiveness Results. At the Morningside Academy in Seattle, and at Malcolm X College in Chicago, students chart their own performance and aim at doubling their performance each week. The results of these x2 learning aims combined with a curriculum designed to support such rapid learning permit the school to give a money back guarantee if students do not gain two grade levels in their subject matter each year (Johnson & Layng, 1992).

Standard Efficiency Results. Standard Change Charts permit comparing effects across departments and years. Workers do not have to stop and figure out each new chart. Precision Teachers have used Standard Change Charts (SCC) since 1970. A 1998 precision teacher or child in her regular elementary classroom can instantly read an SCC chart from a special education Montesori student that was made in 1970.

Standard Change Charts save precious time because they can be read in one minute. Regular standard chart sharings are held every year at the annual conference of the Association for Behavior Analysis (ABA). For over ten years each chart sharer was given two minutes at the overhead projector to present their chart. In 1998 the time was reduced to one minute each with no real loss in audience comprehension. One minute is enough time to read a chart because all in the room are familiar with the standard chart slopes and conventions.

Standard Statistical Results. Standard Change Charts make the up bounces in data equal the down bounces, because frequencies bounce proportionally. Statisticians call this "normalizing the variance." Standard Change Charts also make the total bounce the same size at low frequencies as at high frequencies. Statisticians call this, "equalizing the variance."

Standard Change Charts straighten out the concave upward performance change curves that are always seen with performance changes on add scale charts. These straight lines make it easy to project and see where the performance will end up.

COMPARING STANDARD WITH FILL THE FRAME CHARTS

Many statisticians and managers of management information systems think when they see a Standard Change Chart that it is merely a logarithmic chart of the data. This is not true. To make this difference clear I have prepared six standard charts together with a fill the frame add chart and a fill the frame multiply (Log10) chart of the same performance data time series. Comparing these different chart views shows how Standard Change Charts™ make it easy to read changes from performance improvement charts.

The StatView™ statistical program for both Mac and PC permits setting the exact chart frame size and selecting the axis type and ranges. StatView™ is a trade mark of Abacus Concepts, Inc., Berkeley CA. Obtain further information about StatView™ from www.sas.com. This program also permits making many different charts and views of a data set without having to reenter the data. Therefore, a Standard Change Chart, a fill the frame add scale, and a fill the frame multiply scale chart for each data set can be made. StatView™ was used to make the following figures. Zero Brothers Software at zerobros@aol.com has an online course that teaches how to make standard change charts using StatView™ for both PC and Mac platforms.

Figure 11.1. Dane Learns Cards to Fluency on Daily Chart. Here appear three charts of Dane learning SAFMEDS cards by practicing at least one minute each day, SAFMEDS means, "Say All Fast a Minute Each Day Shuffled." This name was coined to make sure learners practiced saying the whole stack of 70 to 100 cards, practiced at 50 per minute, practiced daily, and shuffled the deck after each practice (Graf, 1994).

The top view shows a Standard Change Chart with its fan of nine easy to recall change factors in the box to its right. To judge the amount of change in a plotted line of data, estimate the change factor using the fan blade as guides. Note that Dane's hits accelerated at x2 per week up to about 80 per minute. Then his hits multiplied at less than x1.4 per week up to his fluency of over 110 per minute. Dane's misses divided at about 16 per week going from 11 to 0 in one week. The misses then bounced from 1 to 0 per minute. About 15 days he practiced more than one one-minute timing each day in his attempts to get over 100 hits per minute. For a discussion of fluency and its advantages read Binder 1990, 1996.

Look at the fill the frame add chart in the middle frame. The counts per minute can be read clearly. But, the learning slope factors can not be read without calculating from the counts because there is no change factor fan and the one at the top right only applies to standard change charts.

Looking at the fill the frame multiply chart in the bottom frame does not help either.

Figure 11.1. Dane Learns Cards to Fluency.

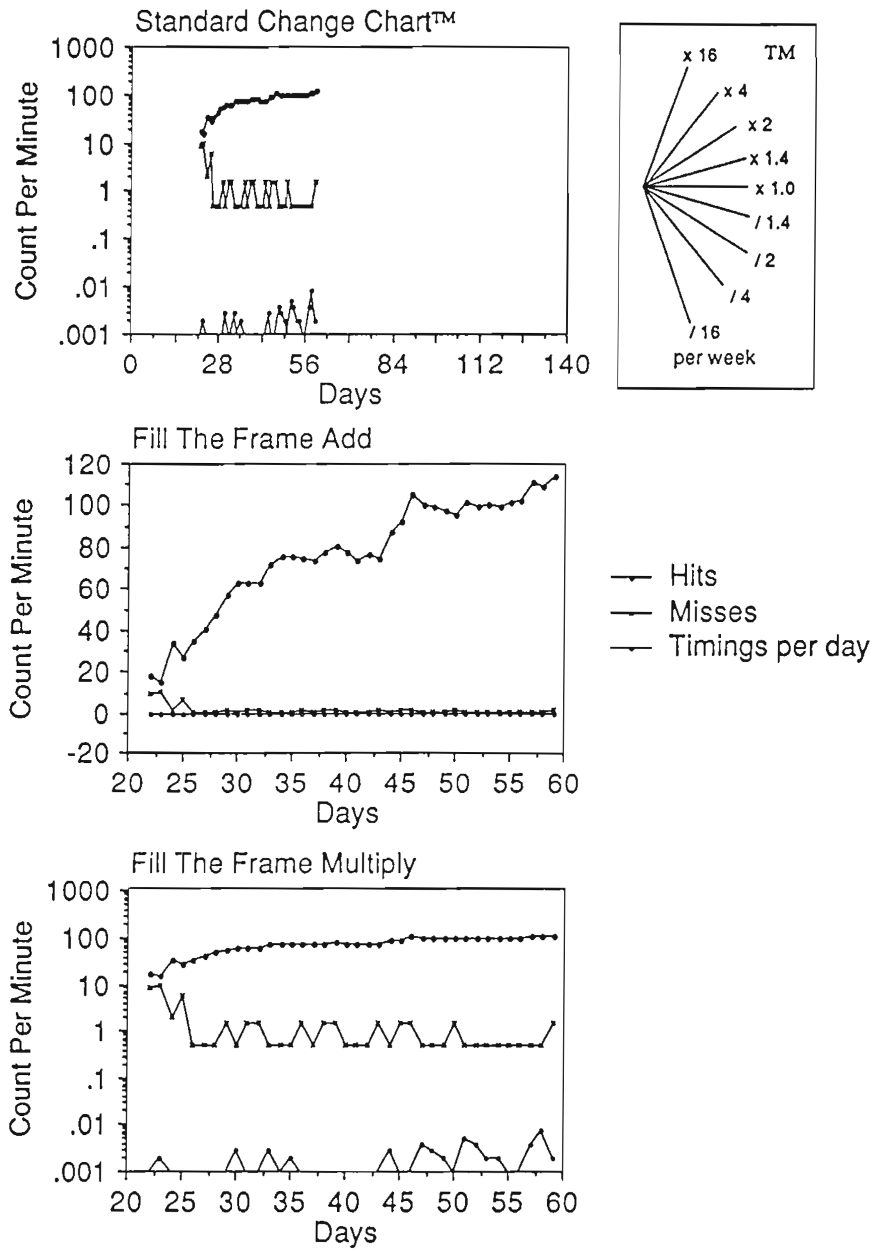


Figure 11.2. Davis Learns Cards Not Fluent on Daily Chart. Here appear the three charts of Davis learning SAFMEDS cards using the same cards and following the same instructions as Dane had received. Both Dane and Davis were students in Graf's 1997 Psychology Class at Youngstown State University. Note from the top Standard Change Chart that Davis' hits accelerated at a little less than x1.4 per week for about 75 days, Then for about 20 days they leveled off at 25 to 28 per minute with no acceleration (x1 per week). Note also that Davis' misses bounced along at 1 to 4 per minute. He only practiced one one-minute timing a day throughout. Perhaps Davis did not practice more because he was happy with the C grade this frequency would earn him.

Note that in comparing Dane's and Davis' fill the frame add charts not much difference is seen without reading the numbers on the charts vertical axes. Their fill the frame multiply charts also do not look very different without reading the axis numbers.

Figure 11.2. Davis Learns Cards - Not Fluent.

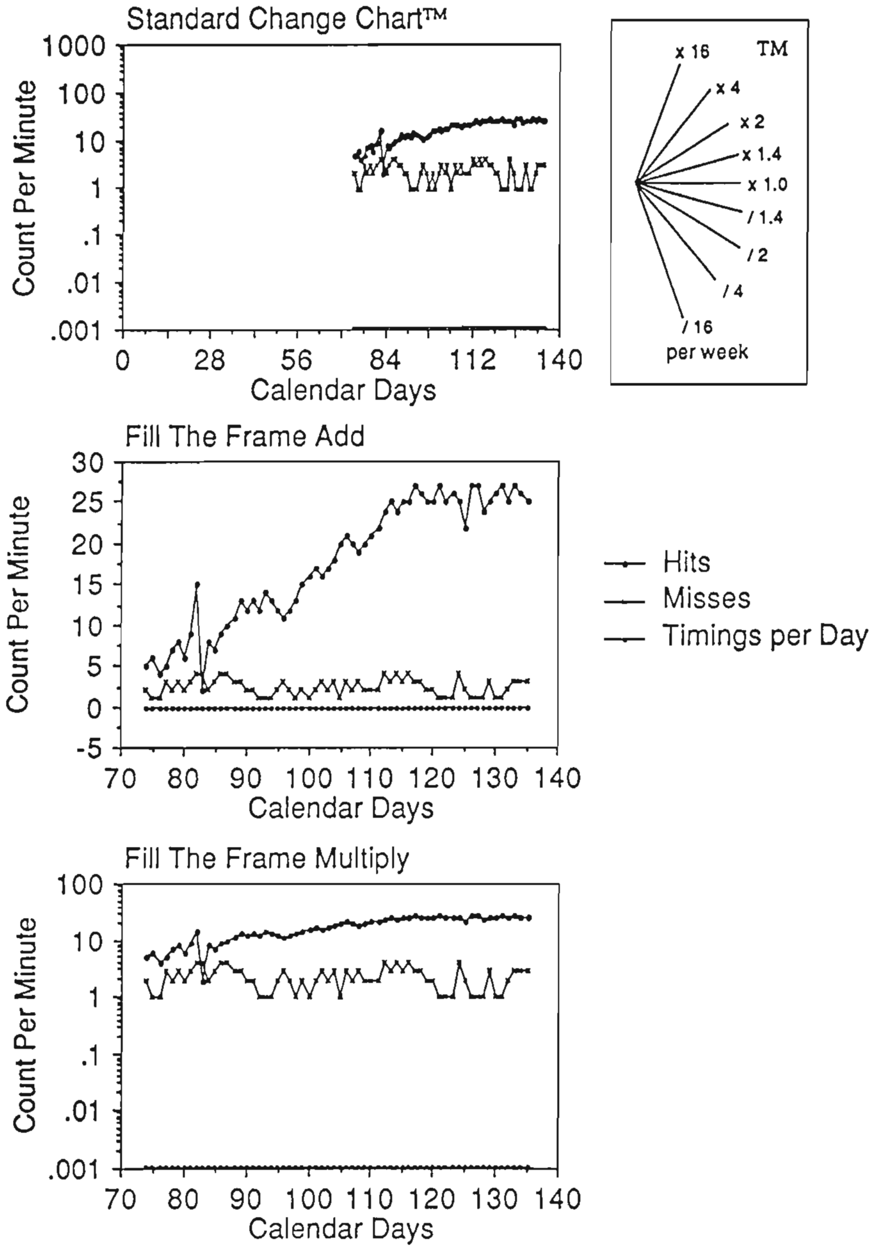


Figure 11.3. Quality in Reebok Plant on Weekly Chart. These three views display the weekly frequency of shoes that conform to specifications and those with defects produced by the Wongpaitoon Footwear Company (WFC) in Thailand. At the start the company was fifth in quality among 12 Reebok manufacturers in Thailand. After performance management it was first in quality among 30 Reebok manufacturers worldwide (Sulzer-Azaroff & Harshbarger, 1995).

Note that the change factor fan blades and values in the box in the upper right of Figure 11.3 are the same as the ones in Figures 11.1 and 11.2, except that Figure 11.3's fan says per month at the bottom, where Figure 11.2's fan said per week at the bottom. The angles and values of the blades are the same, but on the daily chart the change factors are per week, and on the weekly chart the change factors are per month

Look at the Standard Change Chart in the top left frame, and see that conforming shoes that met specifications were not accelerating. The number of defective shoes decelerated at a little less than divide by 1.4 each month for about 20 weeks. Because the deceleration in defects had not leveled off, the program could have continued and achieved even higher levels of quality.

Looking only at the fill the frame add chart in the center of Figure 11.3 would be very reassuring because it looks like 98 out of 100 shoes conform. The quality goal would appear to be reached, and performance improvement efforts might stop.

However, look back at the Standard Change Chart at the top and see the defects are still decelerating and are still at 2000 per week! Projecting this deceleration suggests that if the program were continued, the defects would be down to 1000 per week by November 1993 and maybe even down to 100 per week by November 1994. This forecasting is the great advantage of straight-line projection.

Figure 11.3. Quality Improvement in Reebok Shoe Plant.

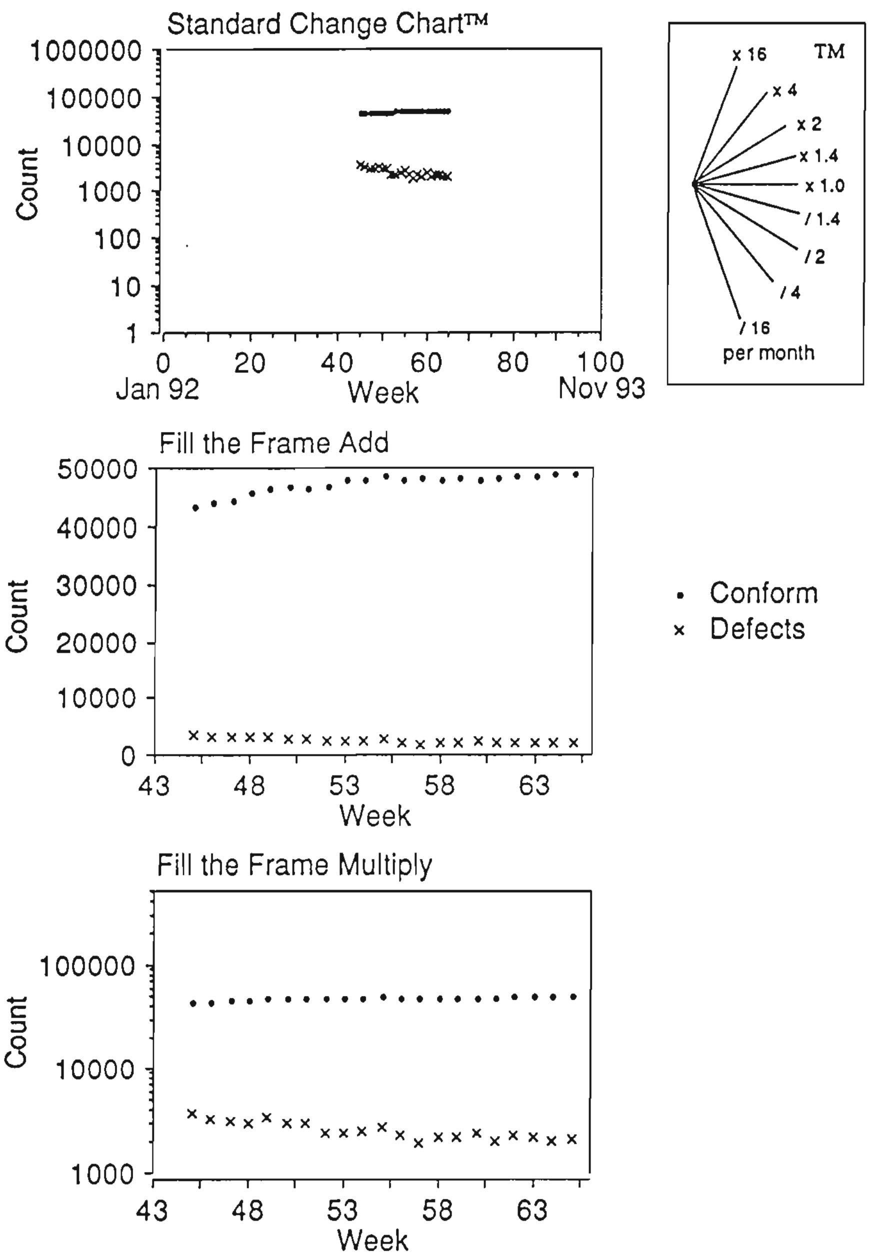


Figure 11.4. Quality at Hewlett-Packard on Monthly Chart. Here see three views of line assembly failures in parts per million (ppm) in the dip soldering process at the Yokagawa Hewlett-Packard (YHP) plant. In 5 years from August 1977 to July 1982 assembly failures were taken from 4000 to 3 ppm (Imai, 1997, pp. 41-43).

Note that the change factor fan blades in the upper right box are the same as those in figures, 11.1, 11.2, and 11.3, except here the change factors are per 6 month periods. The charts shifted from daily to weekly to monthly data with no new angles or factors to learn. Just remember that on the monthly chart the change factors are for 6 month periods.

The top view once again shows the failures on a Standard Change Chart. Note that the failures divide by about 4 every 6 months up to 60 months, the start of 1980. Their deceleration line best fits the /4 blade on the change factor fan. The failures were brought down to 40 parts per million. During this first phase YHP improved working standards, collected and analyzed defect data, introduced process control jigs, trained workers, encouraged quality control circles, and reduced worker's careless mistakes. The jumps in the deceleration are caused by beginning one of these actions.

Still in the top view find the second phase in failure deceleration as the line turns from divide by 4 up to divide by 1.4 every 6 months from the 60th month to the 90th month in July 1982. In this second phase, YHP applied new technologies, revised engineering standards, improved PC board designs and production layout, and added just-in-time concepts. These procedures decreased YHP assembly line failures to only 3 parts per million.

Looking at the fill the frame add chart in the center, note the decrease in failures up to about month 60, but the curve is concave upwards and cannot be projected. The chart shows almost no decrease from the 60th to the 90th month. The add scale hides the valuable deceleration in failures during the second phase of the program. If the add scale were used to track improvement it would look like no further improvement could occur and no further attempts to improve quality would be taken.

The fill the frame multiply chart at the bottom of Figure 11.4 shows fairly straight line deceleration but the turn up at the 60th month is not as clear as in the standard chart at the top. Also, the deceleration factors can not be easily read because no change factor fan exists for a fill the frame chart.

Compare the failure reduction of divide by 1.4 per month at the Wangpaitoon Footware Company (WFC) in figure 11.3 with the divide by 4 per 6 months assembly line failure at Yokagawa Hewlett-Packard (YHP) in figure 11.4, by doing some simple arithmetic. Divide by 1.4 per month must be multiplied by itself six times to find out how big a division factor it would be if carried on for 6 months. This comes out to be divide by 7.5 per 6 months. Divide WFC's /7.5 by YHP's /4 per 6 months and get 1.9. This means WFC's Reebok defect deceleration was 1.9 times more effective than the first phase of YHP's Assembly line failure deceleration. Once again, maybe WFC Reebok stopped their quality quest too soon.

Figure 11.4. Quality Improvement at Hewlett-Packard.

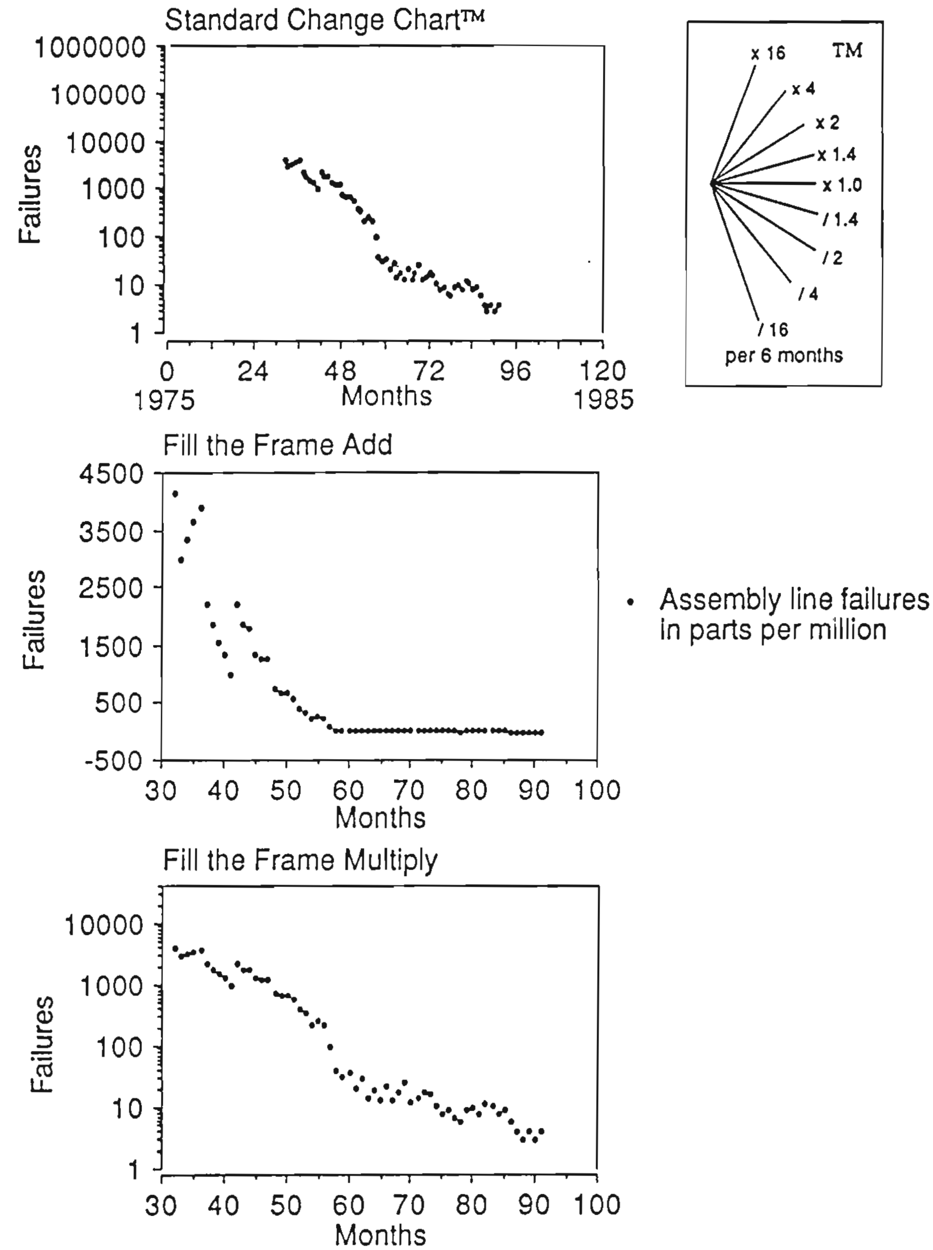


Figure 11.5. Sexual Harassment Prevention on Monthly Chart. These three views show the number of formal internal complaints of sexual harassment in Healthcare, Inc., a large hospital chain (Hill & Phillips, 1997). A sexual harassment prevention workshop conducted during the month of October is indicated by the arrow on each chart view.

The standard change factor fan for monthly charts is in the upper right box. The Standard Change Chart at the top shows the formal internal complaints multiplying by less than the x1.4 fan blade and more than the x1.0 blade at an estimated factor of about times 1.1 every 6 months. After the prevention workshop the complaints decelerated at divide by 2 with a slight turn up at the end. The turn up means the effect may have worn off in one year and they should probably have conducted another prevention workshop. Also, the high turnover in healthcare staff gives another reason to repeat this effective workshop. Although not as steep as the divide by 4 per 6 months of failures by Yokagawa Hewlett-Packard shown in Figure 11.4, this harassment deceleration of divide by 2 is half as big and not so bad for a single prevention workshop.

The deceleration in formal internal harassment complaints shown in the fill the frame add chart in the middle, and the fill the frame multiply chart at the bottom show more detail and exaggerate the workshop effect, but make it almost impossible to compute the change factors. The fill the frame charts also make it look like the harassment deceleration was as steep as the assembly line failure deceleration in Figure 11.4.

Figure 11.5. Sexual Harassment - Formal Complaints.

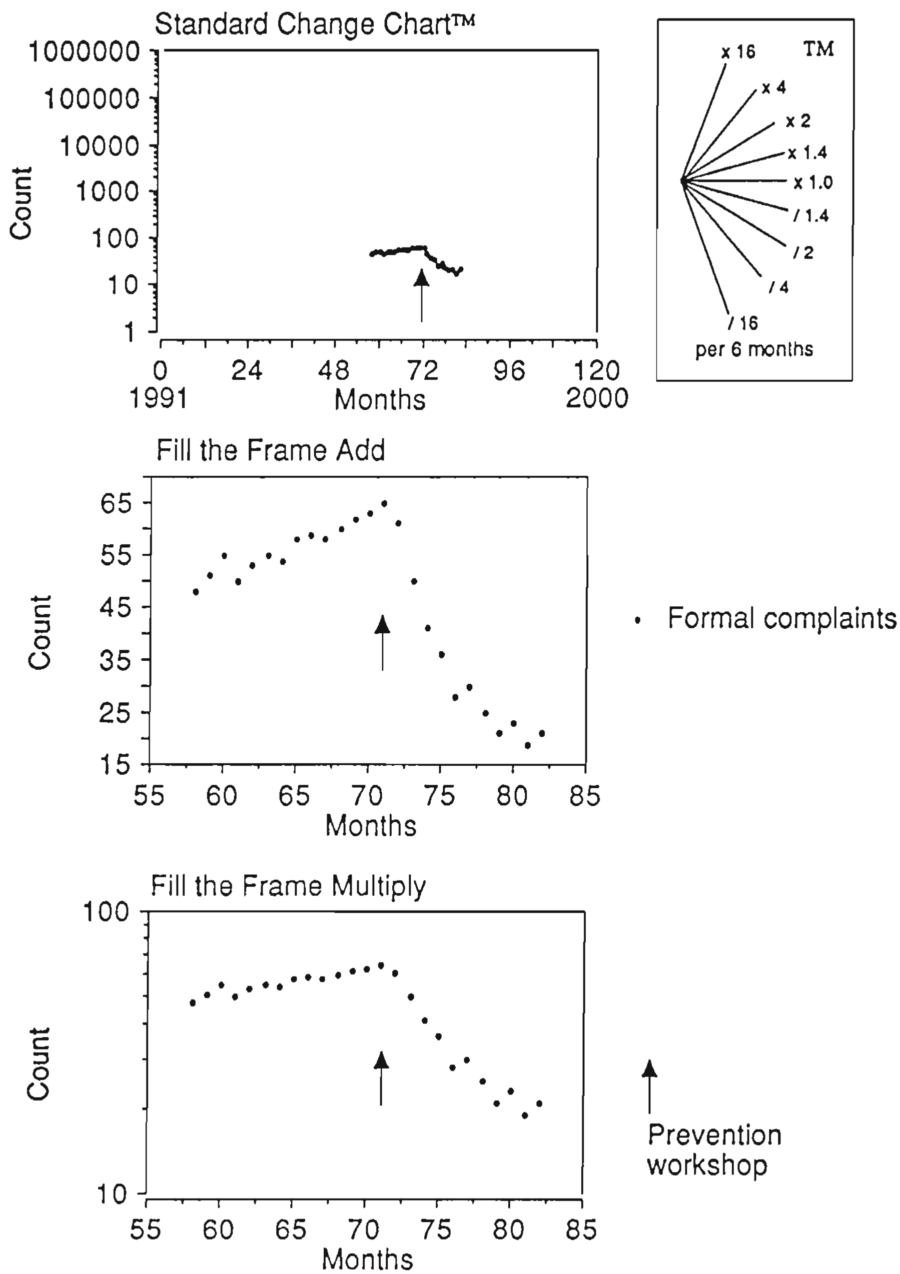


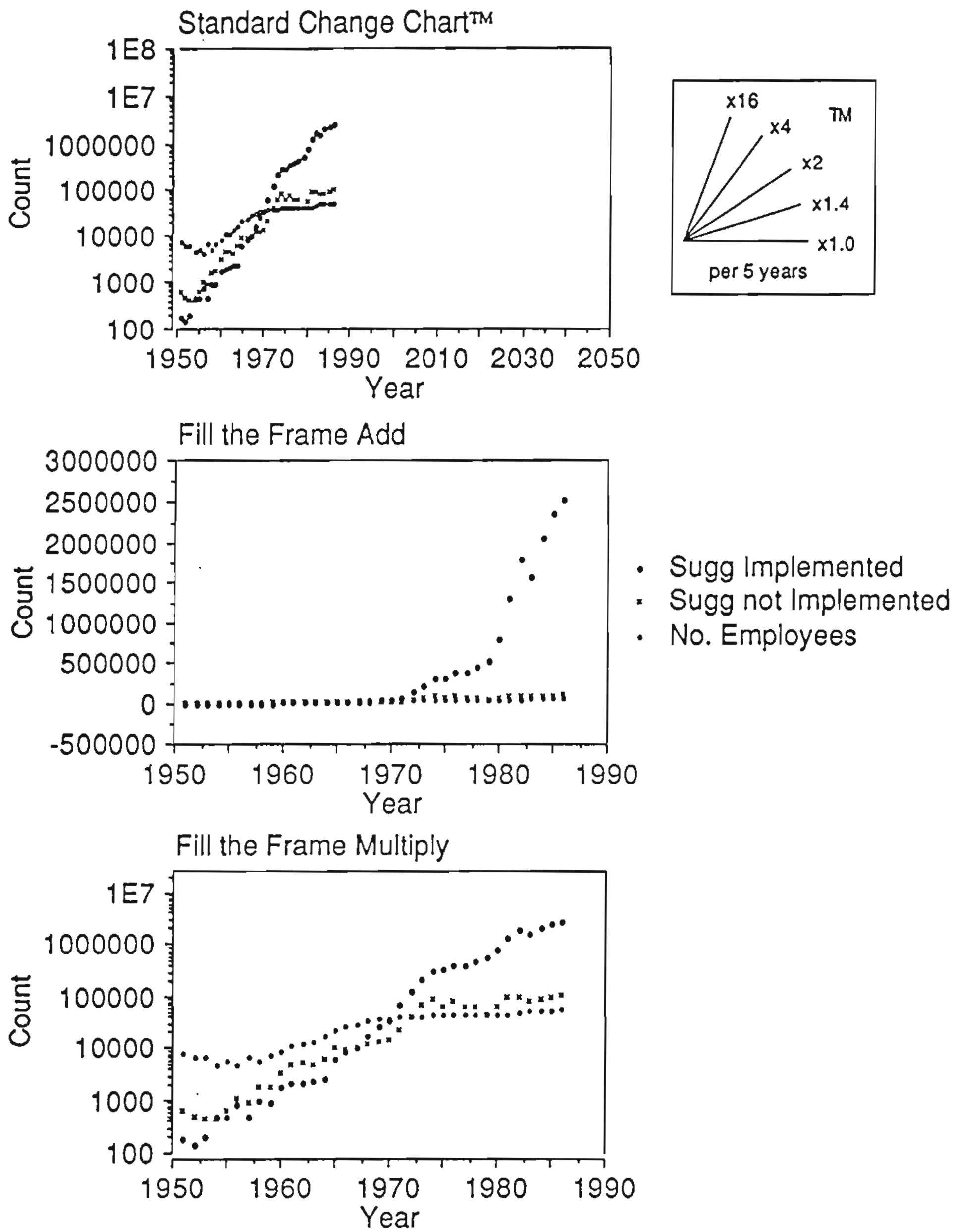
Figure 11.6. Toyota Employee Suggestions on Yearly Chart. Here the three views show the acceleration in employee suggestions implemented and those not implemented each year along with the number of employees at Toyota. Over the course of 35 years from 1951 to 1986 Toyota's implemented suggestions went from 181 per year to 2,542,762. The not implemented suggestions went from 608 to 105,948. The number of employees grew from 7,890 to 55,529.

The change factors fan in the upper right box is now per 5 years for the yearly standard change chart. The yearly Standard Change Chart in the upper frame extends 100 years from 1950 to 2050. Note that the implemented suggestions have fairly consistently multiplied by 4 every 5 years. From 1951 to 1972 the not implemented suggestions multiplied at about the same factor as the implemented suggestions, times 4 every 5 years. Note that the not implemented started out in the first few years above the implemented, Since 1972 the not implemented suggestions have been below the implemented and are multiplying at only about 1.1 very 5 years. This is the same factor as the recent employee growth.

As usual, the fill the frame add chart in the middle view gives us little information on the lower early values hidden by the increases since 1970.

Also, as usual, the fill the frame multiply chart in the bottom view shows an expanded view of what appears in the standard change chart. But, it does this at the expense of preventing familiar standard change factors which immediately tell the size of the growth. It also gives no room for projecting the growth lines to future values.

Figure 11.6. Toyota Employee Suggestions.



PERFORMANCE TRACKING JOB AID

The most useful closing is a job aid to help in tracking performance. Remember KISSING when choosing what to count and chart for impact charts. Chart the daily or weekly impact of performance improvement procedures while using them, and improve procedures while working. Do not be trapped by procedure details chosen at the project start, but focus and refine your methods by tracking their results in action.

Step	Action
1	Determine business issues and needs from focus groups.
2	Locate current business records of daily or weekly frequencies that would track these needs.
3	Reclaim original numbers if record is a percent or cooked data.
4	Have workers tally at their work stations to create records for prime business needs not routinely recorded by your client.
5	Pinpoint a quality pair - something to do more often and something to do less often for each need. Pairs make quality and accuracy.
6	Pinpoint accomplishment pairs that pass Gilbert's "Leave-it Test "
7	Pinpoint behavior pairs that pass Lindsley's "Dead Man Test."
8	Chart daily or weekly frequency pairs on Standard Change Charts.
9	Chart percent records on add scales if business has recorded them.
10	Enter numbers in a spread sheet on a laptop for permanent storage by someone from records office who attends chart share meetings.
11	Post charts prominently in workplace where workers will see them in passing several times each day.
12	Post several copies of same chart if team members use different locations (one outside men's room another outside women's room).
13	Have workers chart data and supervisor check charting accuracy.
14	Rotate charters guaranteeing that all can chart and read the charts.
15	Chart each frequency when it is reported and share all charts (both frequencies and percent) with workers and supervisors.
16	Share charts at start of each day or week rather than at end.
17	Ask all workers for possible causes of all very good or bad days.
18	Act at once upon employee suggestions to improve performance based on charted good or bad days.
19	Mark the change date on the chart with the symbols for the new procedure when it is started.
20	Discuss changes in performance produced by procedure changes at the next morning chart share.
21	Act today on new additions or corrections to the procedure decided upon during the morning chart share.
22	Continue changing to improve through the end of your project.

A time tested Precision Teaching slogan from the 1970's, provides a fitting close: **"Care Enough to Chart."**

REFERENCES

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

Binder, C. (1996) Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19(*2), 163-198.

Brethower, D., & Smalley, K. (1998). *Performance-based instruction: Linking training to business results.* San Francisco: Jossey-Bass Pfeiffer

Daniels, A. C. (1994). *Bringing out the best in people.* New York, N.Y. McGraw-Hill, Inc. 208 Pages.

Dean, P. J. (Ed.). (1994). *Performance engineering at work.* Batavia, IL: International Board of Standards for Training, Performance, & Instruction.

Dickelman, G. J. (1998). Defending human atributes in the pursuit of performance-centered design: An interview with Donald A. Norman. *Performance Improvement, 37*(4), 34-37.

Esque, T. J., & Patterson. P. A. (Eds.). (1998). *Getting results: Case studies in performance imnprovement.* Amherst, MA: HRD Press. Inc. & International Society for Performance Improvement.

Fuller, J. (1997). *Managing performance improvement projects.* Washington, D. C.: International Society for Performance Improvement: San Francisco: Jossey-Bass/Pfeiffer.

Graf, S. A. (1994). *How to develop, produce, and use SAFMEDS in education and training.* Poland, OH: Zero Brothers Software.

Hill, D., & Phillips, J. J. (1997). Healthcare, Inc. In J. J. Phillips (Ed.), *In action: Measuring return on investment, Vol. 2* Alexandria, VA: American Society for Training and Development.

Holzschuh, R. D. (1966). *Superiority of rate correct over percent correct* Unpublished manuscript, Educational Research, University of Kansas Medical Center.

Huff, D. (1954). *How to lie with statistics.* New York: W. W. Norton.

Imai, M. (1997). *Gemba Kaizen: A commonsense, low cost approach to management.* New York: McGraw-Hill.

Industry Report. (1996). *Training*, 33(10).

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

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler Publishers.

Lindsley, O. R. (1971). Precision Teaching in perspective: An interview with Ogden R. Lindsley, Ann Duncan interviewer. *Teaching Exceptional Children, 3*(3), 114-119.

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

Lindsley, O. R. (1994). *Dangers of percent and how to avoid them.* Lawrence, KS: Behavior Research Company.

Lindsley, O. R. (1995). Do, Don't, How; and Did, Didn't, Why? *Performance & Instruction, 34*(2), 23-27.

Lindsley, O. R. (1997a). Performance is easy to monitor and hard to measure. In R. Kaufman, S. Thiagarajan, & P. MacGillis (Eds.). *The guidebook for performance improvement: Working with individuals and organizations.* (pp.519-559). San Franciso, CA: Pfeiffer, Jossey-Bass.

Lindsley, O. R. (1997b). Precise instructional design.: Guidelines from Precision Teaching. In C. R. Dills & A J. Romiszowski (Eds), *Instructional development paradigms* (pp. 537-554). Englewood Cliffs, NJ: Educational Technology Publications.

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

MEASURIT. (1996) *Tracking training impact.* Shawnee Mission, KS: MEASURIT.

Mager, R. F. (1986). *The how to write a book book.* Carefree, AZ: Mager Associates, Inc.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Parker, M., & Leinhardt, G. (1995). Percent: A privileged proportion. *Review of Educational Research, 65*(4), 421-481.

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

Phlllips, J. J. (Ed.). (1994) *In action: Measuring return on investment, Vol.1.* Alexandria, VA: American Society for Training and Development.

Phlllips, J. J. (Ed.). (1997) *In action: Measuring return on investment, Vol.2.* Alexandria, VA: American Society for Training and Development.

Ries, A. & Trout, J. (1993). *The 22 immutable laws of change: Violate them at your own risk!* New York: HarperCollins Publishers.

Schmid, C. F. (1954). *Handbook of graphic presentation.* New York: Ronald Press.

Shewhart. W. A. (1939). *Statistical methods from the viewpoint of quality control.* Washngrton, DC: Graduate School, Department of Agriculture; reprint edition, 1986, New York: Dover.
Skinner, B. F. (1938). *The behavior of organisms.* New York: Appleton-Century-Crofts.
Skinner, B. F. (1950). Lecture in *Pro-Seminar in General Psychology*, Psychology 201a, Harvard University, 27 November, 4 and 11 December 1950
Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis.* New York: Appleton-Century-Crofts.
Spitzer, D. R. (1998, May). *Super-Evaluation: A new paradigm for increasing the results you get from training.* Paper presented at the meeting of the Kansas City Chapter of The International Society for Performance Improvement.
Stolovitch, H. D. &, Keeps, E. J. (1992). (Eds). *Handbook of human performance technology.* San Francisco, Jossey-Bass.
Sulzer-Azaroff, B., & Harshbarger, D. (1995). Putting fear to flight. *Quality Progress, 28*(12), 61-65.
Tosti, D. T. (1978). Formative feedback. *NSPI Journal, 19*, 19-21.

BRIEF BIOGRAPHY OF AUTHOR

Ogden R. Lindsley is President and Founder of Behavior Research Company, provider of performance tracking and change charting materials and systems to schools and corporations since 1960. A 1957 Skinner Ph.D. from Harvard, he founded the first human operant laboratory, coined "Behavior Therapy," was President of the Association for Behavior Analysis, and is a Kansas University Professor Emeritus. He developed Precision Teaching and Standard Change Charting. Ogden received the Thomas F. Gilbert Distinguished Professional Achievement Award from the International Society for Performance Improvement in 1998. Website at WWW.onlearn.com/lindsleybio.html Reach Og at OLindsley@aol.com.