Measurement: A Few Important Ideas

by Carl Binder, PhD

Measure: 1. the dimensions, quantity, or capacity of something as ascertained by measuring; 2. a reference standard or sample used for the quantitative comparison of properties. *(The American Heritage Dictionary, Second College Edition*).

Lindsley (1999) and others have observed how few International Society for Performance Improvement (ISPI) publications or presentations contain measures of performance or business results, despite our claims to produce measurable improvements. Many of us came to the field of human performance technology (HPT) without any background in measurement, scientific analysis, or other prerequisites for good measurement practice. A growing number of our colleagues who have taken statistics courses in graduate programs or who have been influenced by publications in educational research seem to believe that we need advanced math skills and statistical packages to conduct good performance measurement. For all these reasons and more, many of us are uncomfortable with performance measurement and have not incorporated it as a regular part of our day-to-day practice.

In its most rudimentary form, measurement involves counting and comparing. We imagine cave dwellers lining up stones in one-to-one correspondence with animal carcasses to represent the success of their hunting efforts, and perhaps comparing piles of stones representing different expeditions or hunting parties. We've heard the history of the *foot* measure based on the length of a king's body part—an early example of standardization. Counting volume in *barrels* and weight in *stones* similarly conjures up images of primitive standards resulting in the ability to compare and share measures of various types by counting agreed-on units. Fundamentally, measurement boils down to these simple notions of units and standards.

This article presents a few important ideas, some principles, and a collection of examples intended to clarify thinking about performance measurement. By no means does it attempt to present a comprehensive view of measurement in HPT. Instead, it emphasizes three key points:

- The foundation of performance measurement is to identify *countable* behaviors, accomplishments, and business results.

- Performance always occurs in *time*, and therefore performance measurement should not ignore the time dimension.
- Representing differences and changes as multiplying or dividing factors is preferable to using percentages, in part because a given percentage increase is not equivalent to the same percentage decrease.

By clarifying these basics, I hope to demystify measurement and encourage practitioners to measure and think about measuring performance and learning in simpler, more meaningful ways. I want to discourage use of measures that lack standard units, and encourage an approach consistent with that of the natural sciences and engineering (Binder, 1995). The upshot, I hope, will be to prompt more direct measurement of performance and greater sharing and comparison of results.

Reasons for Measurement

A good place to start is to consider why we measure. There are essentially three reasons for measurement in the science and practice of performance improvement: validation, accountability, and decisionmaking. Each of these reasons has implications for how we measure and how we use the products of the measurement process.

Validation

Most research studies, journal articles, and reports focus on validation of methods, procedures, or programs. They present measures to demonstrate that a particular intervention or approach produced the desired effect (or failed to do so) or worked better than another. But as any experienced performance improvement practitioner can attest, simply repeating a supposedly validated study in a new setting or different case is no guarantee that it will work. Given natural variability in the real world, whether a program or intervention worked in one setting does not always predict whether it will work in another, or even in the same setting at a different time. Consequently, for the practitioner, *validation data are not sufficient.* One must continue to measure in any new situation to determine whether the intervention actually works in that case.

Accountability

I once described accountability as a reason for measurement. The client's immediate response was, "Oh, you mean cover-your-a** data!" I agreed, and we went on to discuss the fact that many organizations require measures of process or outcomes for administrative purposes. They want to justify or document whatever program, process, or intervention they are implementing. Often, when we ask for measures, clients give us stacks of spreadsheets or graphic displays that show, for certain, that an intervention happened and that someone was monitoring it. Unfortunately, these data are usually filed away and have little impact beyond the satisfaction of requirements. They might serve to justify budgets, but they are often measures of process, not results—and they usually do not come in a form that easily supports timely decisions about what to do next. In training programs, for example, final test scores are often reported to managers in a way that has little effect other than to show that the training organization has been busy and that trainees can pass tests. In rare cases, training personnel conduct item analysis to identify content for which they should improve training; this is a form of decisionmaking. But, for the most part, accountability measurement simply meets administrative requirements for documentation.

Decisionmaking

When airplane pilots check their instruments, they use data to make decisions and course corrections. When business executives check monthly financial reports, they use data to adjust strategies, tactics, budget allocations, and the like. For scientists, technologists, and practitioners in many fields, a principal purpose for measurement is to support decisions about what to do next, how to adjust procedures, or when to make changes. By measuring in a way that will support good decisions, it is usually also possible to document accountability and to validate any interventions that prove generally successful. So from this perspective, *decisionmaking is the highest purpose for measurement*, and it generally subsumes the other two purposes.

One of our senior colleagues, a prominent cognitive theoretician, once remarked about measurement that "Skinner had it right on measurement, whatever else he might have gotten wrong." To understand that comment, it is important to know the fundamentals of Skinner's measurement system (Skinner, 1938).

In the laboratories where he developed his experimental analysis of behavior and in which he discovered basic principles of reinforcement schedules, discrimination learning, and so many other foundations of contemporary behavior science and technology, Skinner counted the responses of his subjects using automatic recording devices. He also timed the periods during which the responses occurred. He monitored the frequency (count per minute) of those responses and watched for variations or changes in frequency on continuously drawn graphs that allowed him to make timely decisions about how to change experimental programs or conditions. He developed a sensitive measurement system that supported decisionmaking while at the same time allowing him to later summarize the data to validate discoveries.

Performance improvement professionals, no matter what their theoretical orientation, will optimize the power of their interventions to the extent they can emulate certain basic features of Skinner's measurement system.

Performance Counts

Returning to my initial example of hunters quantifying their success with stone counters, it's important to notice that most measurement that would pass criteria for objectivity in the natural sciences comes down to counting. Once we have identified a behavior or performance to measure, we decide on a countable unit. This was the essence of Skinner's measurement system: He figured out how to count the critical effects or "accomplishments" of subjects in experimental situations (e.g., switch closures produced by pressing keys on a laboratory apparatus) and monitored those counts continuously over time (Gilbert, 1996).

GILBERT'S REQUIREMENTS	CORRESPONDING COUNTABLE UNITS
Quality	
Accuracy	Count of accurate items Count of inaccurate items
Class	Count of items in each category, rating, or class
Novelty	Count of items defined as novel, in separate categories if desired (e.g., new method, new product, new/better outcome, etc.)
Quantity (or Productivity)	
Rate	Count of any behavior or accomplishment per unit of time (minute, hour, day, week, etc.)
Timeliness	Count of timely events or items Count of untimely events or items
Volume	Count of items (as in "sales volume") OR Count of volume in units (e.g., liters, cubic yards)
Cost	
Labor cost	Count of dollars spent on labor, by category
Material cost	Count of dollars spent on materials, by category
Management support cost	Count of dollars spent on management support, by category

Figure 1. Countable Units Corresponding to Gilbert's Requirements.

In financial analyses, we count dollars spent and earned. In business operations, we count sales calls, website hits, widgets, positive customer comments, pages produced, and many other identifiable units. In training, we often count correct and incorrect responses or answers to test items. In each case we define or pinpoint a unit and count it. While we might derive a variety of percentages, ratios, or formulas based on these counts in subsequent analyses, the foundation is the countable units that best reflect our objectives.

Countable Units for Gilbert's Requirements

Thomas F. Gilbert, regarded by many as the father of HPT, listed nine "requirements" or types of criteria that one might evaluate in efforts to define or improve human performance (Gilbert, 1996, p. 45). While he did not describe how we might turn each of these into countable units, it's simple enough to do the translation, as shown in Figure 1.

I propose that performance measurement needs to include the time dimension to be optimally useful. Rather than only including time with productivity measures, I'd argue that we should also count units of cost and quality over time. As mentioned above, this was another essential ingredient of Skinner's measurement system—counting behavior rates or frequencies of occurrence over time and using count per minute as his basic datum. In organizations, we might use count per hour, per day, per week, per month, or per quarter for different measures. But, for the most part, if we do *not* include the time dimension with our counts, we're likely to make less-effective decisions about the value of performance or the worth of interventions. It is virtually always important to know the period during which the target behavior, accomplishments, or results occurred.

Behavior and performance, because they occur in time, have essential temporal features. Slow, accurate performance is different from fast, accurate performance. A profitable company with low revenues is different from a profitable company with high revenues. Ignoring the time unit—for example, per minute, per hour, per day, per week, per month, or per quarter—removes measurement from the realm of real-world performance.

As a far out example, compare an artist who produces a single world-class painting (reflecting Gilbert's class or novelty requirements) with the performance of a Van Gogh or Picasso. We'd likely value the latter more highly because of how many such paintings they were able to produce over time.

The quantity, quality, and costs associated with human performance and results in an organization are best evaluated using a time dimension, at both macro and micro levels of performance. At the macro level, it's clear that financial results and productivity measures necessarily include time periods. Monthly or quarterly measures of revenues and expenses are typical, while hourly, daily, or weekly measures of productivity are common. At a more micro level, research on behavioral fluency (Binder & Bloom, 1989; Binder, 1996) highlights the enormous practical differences indicated by completion of tasks or tests in different time durations. Being able to perform a task or use knowledge with greater fluency, or automaticity, represents a higher degree of skill that is not reflected in time-less accuracy measures. As an example, one who can immediately select the correct responses in a multiple choice test is said to "know the material" better than one who requires more time for guessing and comparison of test items to produce the same level of accuracy. Knowledge measurement procedures that do *not* include a time dimension are relatively insensitive to huge differences in expertise.

What advice would you give someone on the path to becoming a master in his or her field?

I recommend doing whatever it takes to study and learn personally from individuals whose work you admire. I believe that directly measuring the behavior, performance, and results that you seek to improve is the best way of keeping honest and of advancing your own development and that of your field. I strongly advocate writing and publishing as a regular habit. I suggest taking the time to help others learn and develop, both because teaching is one of the best ways to learn, and because help given to others so often comes back to us in wonderful and unexpected ways.

The Dangers of Percentages and Ratios

Lindsley (1999) has written extensively about problems associated with relying strictly on percentage or ratio calculations to inform decisions about performance. When we calculate a percentage or ratio and then use it as the sole indicator for making decisions, we derive what has been called a "dimensionless quantity" (Johnston & Pennypacker, 1980) *that leaves the original performance measures behind.* That is, we do not know from a percentage the number of opportunities (denominator in the percent calculation) nor the time during which they occurred. And there is a huge difference between, for example, 90% of 10 versus 90% of 1000.

As another illustration of this point, when computing a profit percentage in a financial analysis, dollar units in the numerator and denominator cancel out, leaving a percentage with no units of measurement. While this number may be useful in evaluating the profitability of a company, if we do not also include the original revenue and expense dollar figures, it is impossible to know how many dollars the company is actually making. Financial managers know this—how often have you seen a balance sheet that includes percentages without the original revenue and expense numbers? (Never.) Similarly, when we use percentages to represent performance on a test or proportion of rejects in a manufacturing environment, the percentage figures do not tell us anything about the number of test items completed or the number of widgets manufactured. By canceling out the original counts, we eliminate standard units of measurement, thereby eliminating any direct measure of the performance itself. Using percent-correct figures to assess the results of a training program removes that assessment from the realm of performance. It's no wonder that percent-correct test scores so seldom reliably predict performance levels or the time required for ramp-up to benchmark performance on the job!

The important point here is that calculating and presenting percentages or ratios *without also including the performance counts from which they are derived* can distort our understanding of a given performance and therefore our decisions about how to manage that performance. This is as true in knowledge testing as it is in profit and loss statements. The guiding principle should be to present percentages or ratios as supplementary indicators for the convenience of the reader, in addition to the actual performance counts. In that way, measurement will accurately and directly reflect the levels of behavior or accomplishment that are important to those with a stake in the value of the performance or in the worth of the performance intervention.

Standards

Standards are at the heart of any reliable measurement system. The recent vote-counting debacle in Florida's presidential election brought that point home in a dramatic way. If there is a lack of agreement on what is to be counted, then the count itself is in doubt. In the natural sciences, standard units such as meters, grams, and minutes form the foundation for communication, collaboration, and sharing of results. In economics, units of currency and conversion rates between them serve as standards. In any effort to measure performance, we need to agree about standard units and dimensions.

Once we agree on a countable unit of behavior or accomplishment and a unit of time during which (or *per* which) the unit will be counted, then it is possible to collect a verifiable measure. While this might seem obvious, it is remarkable how few intervention reports include results measures in standard units that allow comparison with results of interventions reported elsewhere.

For example, it is common for journal articles or presentations about research on training to include graphs on which the time scale across the bottom is "sessions." If the measures of correct or incorrect performance accelerate at a given angle over the course of that scale, what can we say about the power of the training method? How rapidly does it change performance? If we don't know exactly when the sessions occurred in standard calendar time, it will be impossible to compare the picture of acceleration on that graph with another graph showing sessions across the bottom that occurred on different days, with or without weekends interspersed, etc. Using standard time scales is thus essential for meaningful evaluation of measured results that occur over time.

As an informal investigation, review any data sets on learning or performance improvement that you can find in books, journals, or project reports. Try to determine whether or not the units of measurement (either what was counted or the time during which it occurred) are standard. You might be discouraged to find an amazing lack of standardization in measurement—a lack that will make it difficult or impossible to compare results or effectiveness across different efforts to achieve similar outcomes.

Measuring Changes in Performance

In addition to measuring levels of performance, it is important in training and performance improvement efforts to measure changes in levels of performance in two ways: as *increments or decrements*, and *as rates of change* (trends) in performance.

Increments or Decrements in Performance

It is traditional to describe increases or decreases in performance either in the absolute value of the change (e.g., dollars per month, sales contacts per week, questions answered correctly per minute) or as a percentage of change from an initial baseline or comparison level. Following the logic of the previous comments about the dangers of percent, one should certainly report the absolute counts. However, for comparing changes across different projects or interventions where the baseline values are different, percentages have the advantage in that they allow one to compare proportional changes independent of the absolute value of the baseline. Unfortunately, it is another of the "dangers of percent" (Lindsley, 1999) that a certain percentage increase does not have the same absolute

PURPOSE AND MEASURES	PROCEDURE AND/OR COMMENTS
To measure learning in a classroom program or module using pre- and post-tests • Count per min of correct responses • Count per min of errors + skips	1. Use multiple-choice tests with more items than an expert can complete in the allotted time (e.g., 4 minutes). Use the same test items for pre- and post-test procedures. 2. Instruct participants to complete as many items correctly as possible in the allotted time, and provide clear start and stop signals ("*Please begin*" and "*Please stop*"). 3. Calculate counts per minute, as well as the accuracy ratio (corrects/errors). 4. Calculate how correct responding multiplied and how errors divided over the course of the training program (e.g., corrects x2.1, errors /4.5). This method will reveal differences in knowledge fluency that cannot be seen in conventional multiple-choice tests because it allows unlimited time and produces only percent-correct data.
To measure how many (and which) new ideas or facts were acquired during a classroom session • Count of ideas/facts jotted down in a brief time interval (one or two minutes) at the start of the session or segment • Count of ideas/facts jotted down at the end in the same duration	1. Introduce a topic by first asking a question: "What do you know about X?" Explain that you will be conducting a brief timing (e.g., 1 minute) and instruct participants to jot down (abbreviate, to save time) as many relevant facts, ideas, etc. as possible in the allotted time. 2. Provide clear start and stop signals. 3. After the timing, count the ideas/facts for each individual, record the counts, and share the ideas/facts on a flip chart or white board, making comments as needed. 4. Repeat the procedure at the end of the segment and compare counts of ideas/facts. Using this procedure, you will begin to learn what multiplier effect to expect from presentations (e.g., x 2.0 ideas or facts per minute). The procedure also provides a great opportunity for active responding, sharing, and discussion —*combining measurement with learning*.
To monitor learning the skill of finding information in a customer service performance support intranet site • Count per minute of answers/bits of information found on the site in response to specific questions, monitored over days	1. Create practice sheets (with several variants) of questions to be answered by locating information on the intranet site. Include space for writing down the answers. 2. Allot a fixed period for measurement each day (e.g., a 5-minute timing). 3. Using different versions of the sheet each day, conduct timings (in addition to any other practice that occurs) to monitor change in the navigation skill. Use clear start and stop signals. 4. Count and graph the count per minute of questions answered correctly each day. This approach provides a *direct* measure of a skill that trainees will use on the job. By measuring daily, the procedure allows you to assess learning (change in performance) and to make program adjustments if individual or group performance is not accelerating sufficiently.

Figure 2. Counting and Timing to Measure Learning.

SHAWMUT BANK

MEASUREMENT	CORRECT PER MIN.	ERR+SKIPS PER MIN.	ACCURACY RATIO	RESPONSE TIME
Pretest	5.75	1.75	3.29	8.00 sec.
Post-test	16.00	.25	64.00	3.69 sec.
Improvement ratio	x2.78	/7.00	x19.51	/2.17

FIRST AMERICAN BANK

MEASUREMENT	CORRECT PER MIN.	ERR+SKIPS PER MIN.	ACCURACY RATIO	RESPONSE TIME
Pretest	4.25	2.25	1.89	9.23 sec.
Post-test	14.75	.50	29.50	3.93 sec.
Improvement ratio	x3.47	/4.50	x15.61	/2.35

Figure 3. Examples of 4-min. Pretest and Post-Test Measures.

value as an equal percentage decrease. For example, if I increase sales from 100 units per month by 20%, I'll produce 120 units per month. But if I then reduce 120 units per month by 20%, the result will be 96 units, not the original 100 units. So percentages can be deceptive if used to quantify both increments and decrements in performance, because the value of a given percentage increase is not equal to the same percentage decrease.

An alternative is to use ratios or multiply/divide factors to quantify increases or decreases. For example, we might say that a performance multiplied by 1.2 or divided by 1.2. If an intervention multiplies a 100-unit performance by 1.2, the result is 120 units. Dividing 120 by 1.2 returns us to the original 100 units. Try this exercise several times with percentages and multiply/divide factors on different baseline numbers to convince yourself that representing increments or decrements in performance with multiplying or dividing factors, unlike using percentages, is *symmetrical*—it yields the same results in both directions. This is an important insight contributed by Lindsley (1999), but not recognized by most people involved in measuring or describing changes in performance.

Trends in Performance

We can also quantify trends in performance using multiplying or dividing factors calculated over standard intervals of time. We can say, for instance, that a performance multiplied by 1.5 per month (x1.5) or divided by 1.5 per month (/1.5). Here again, percentages can be deceptive if one believes that a given percentage trend *up* is comparable to a given percentage trend *down*. Growing by 20% per month is not just the opposite of shrinking by 20% per month. (Try this with a few numerical examples to prove it to yourself. Growing by x1.5 per quarter is exactly the reverse of shrinking by /1.5 per quarter.) *Using multiply/divide factors to quantify changes and differences is simply easier to understand without mistakes.* Lindsley's charting methods allow users to present measures of change in a standard graphic format corresponding to these multiply/divide factors and are worthy of careful study by any serious performance improvement professional (1997, 1999). They make description, analysis, and communication much easier and more foolproof.

This multiply/divide approach to quantifying change in performance is especially useful in any effort to assess the relative power of interventions or methods, since it allows direct comparison of magnitudes of effect. *If HPT as a field is going to advance by accumulating reliable information about effectiveness, it is essential that we use tools that allow rapid comparison, communication, and evaluation based on standard units.*

Examples

One of the benefits of adopting these principles—identifying countable behaviors and accomplishments, using a time dimension whenever possible, and representing differences

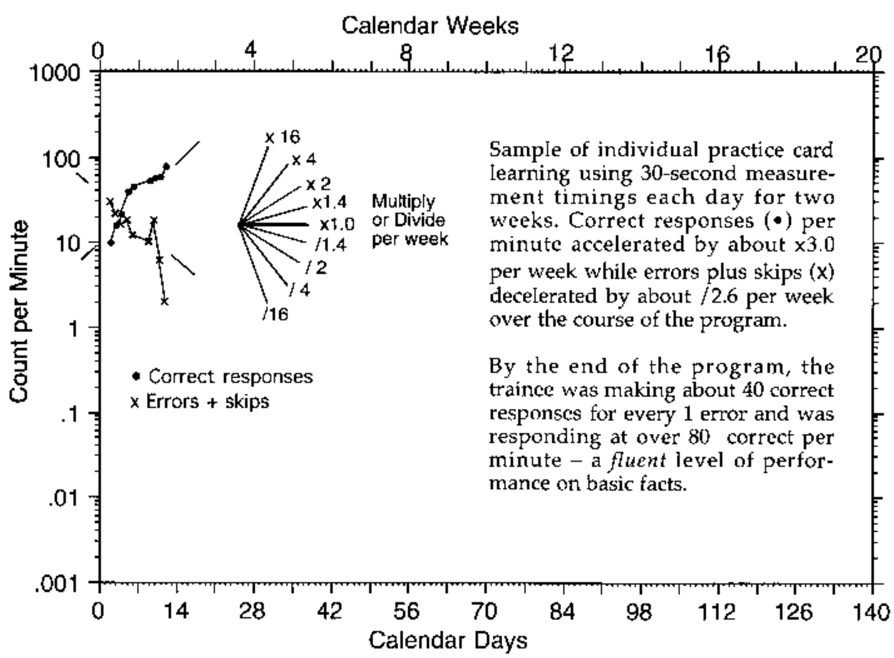

Figure 4. Count-per-Minute Data.

PURPOSE AND MEASURES	PROCEDURE AND/OR COMMENTS
To measure productivity in a customer call center • Count customer calls handled	1. Use automated telephone equipment to count and time customer calls (count per hour). 2. Track performance to obtain ramp-up (trend) and changing ratios between call center benchmark and performance of newly trained representatives. Use the data to evaluate effectiveness of new hire training and coaching programs. Counting and timing performance in both training and on the job allows managers to view results in comparable graphic displays and numerical values.
To measure call quality in a customer call center • Count calls in which representatives exhibit specified behavior(s) (e.g., used the customer's name, asked how else they could help the customer) • Count calls in which representatives do not exhibit the specified behavior(s)	Options for collecting these measures: • Sample calls across multiple representatives, sum for the group (count per day or per week) • Sample multiple calls for individual representatives (count per day or per week) • Arrange for representatives or supervisors to count and record their own behavior (count per day) Note: Counting periods longer than a week provides too few data points for timely decisionmaking and feedback to representatives.
To measure customer contacts by medical device sales representatives • Count office calls on physicians	Collect data from sales representatives' online activity logs (count per week).
To measure effects of equipment maintenance in a factory (partial list) • Count number of scheduled and unscheduled maintenance jobs • Count number of hours between failure for machines	Collect these data from maintenance records (count per month). Summarize data per section of the factory, per machine, per maintenance team, etc.—depending on your specific objectives—and examine both the *range* of time between failure and the average time between failures.
To measure timeliness in a project management environment • Count number of deadlines met • Count number of deadlines missed	Count per week or per month, depending on the number of individuals and projects involved. The shorter the counting period, the more opportunities there are for making decisions or course corrections. Note: By looking at counts, in addition to ratios between them, you can see how the number of deadlines set affects whether they are met—something that you cannot see if you merely summarize the data with a ratio or percentage and leave out the counts.

Figure 5. Counting and Timing to Measure Job Performance.

and changes as multiply/divide factors—is that they can apply in virtually any situation. Figures 2, 3, 5, and 7 present examples illustrating the principles outlined in this article for measuring and presenting data on learning, on-the-job performance, and business results.

Learning

The traditional approach to measuring learning has been to assess accuracy only, using percent-correct scores. Such an approach entirely disconnects learning measurement from performance. The field of precision teaching (Binder & Watkins, 1990; Lindsley, 1997) and the FluencyBuilding™ methodology (Binder & Bloom, 1989) present robust models for using count-per-minute measures to assess progress in learning and coaching programs. Figure 2 presents a few examples based on that work, demonstrating how counting and timing can provide a measurement bridge from the classroom to job performance.

Figure 3 presents some pretest and posttest data, reprinted from Binder and Bloom (1989). Figure 4 is a chart showing count per minute of correct responding accelerating at 3.0/week and errors decelerating at about /2.6 per week for a flashcard practice activity in a customer call center new-hire training program.

On the-Job Performance

Figure 5 offers a few examples of using counting and timing to measure productivity, service quality, and other types of job per-

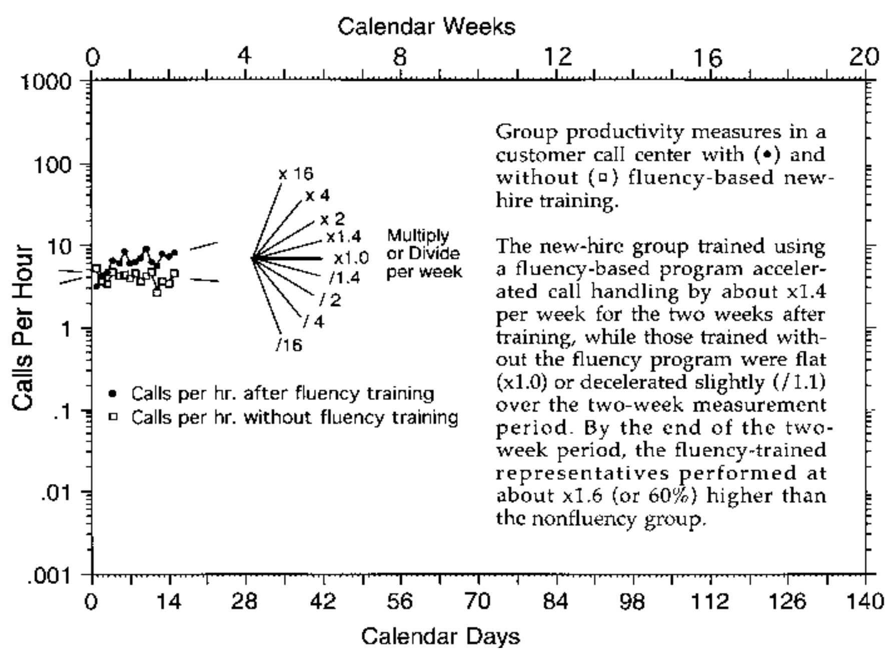

Figure 6. Average Call-Handling Productivity.

PURPOSE AND MEASURES	PROCEDURE AND/OR COMMENTS
To measure sales results • Count proposals submitted • Count units sold • Count dollar value of units sold • Count number of deals/orders	The usual counting period for these data is per month, with a quarterly summary. Monthly data typically support field management decisions, while quarterly data support regional and national Director/VP decisions. With these counts, it also possible to derive the ratio of closed deals to submitted proposals, units per deal, and dollar value per deal, which can be useful indicators, comparing across sales people, sales groups, division, etc.
To measure customer satisfaction • Count number of survey items completed • Count number of survey items (by item, if desired) that customers rated as "very good" or "excellent" • Count number of positive comments on customer hot line • Count number of negative comments on customer hot line	Using counts of responses in different categories (Gilbert's class requirement) retains numerical values that have direct, objective value (e.g., "120 customers out of 180 rated us as 'very good' or 'excellent' this month"). Counting provides numbers that you can add, subtract, multiply, and divide meaningfully. This is in contrast to the widely used procedure of calculating averages from rating scales (e.g., "We averaged 3.75 on a scale of 5"). Such numbers lack an objective reference and do not directly quantify behavior, performance, or tangible results. Technically, it is not meaningful to calculate average ratings based on rating scales because such scales provide *ordinal* data representing *a hierarchy* of values but not equal quantitative intervals (Stevens, 1958).
To measure success in a market • Estimate count of prospective customers • Count actual number of customers	These counts allow calculation of the percent or ratio of market penetration, a traditional marketing success indicator. Again, however, it's best not to simply report ratios or percentages, but also the counts themselves in order to understand the absolute size of the market(s). If there are multiple market segments, market size information is crucial for most decisions.
To measure research and development success • Count number of projects for developing products or patents • Count number of products or R&D projects at various stages in the development process • Count number of approved patents • Count number of approved/new products	These measures are probably best summarized per year, although quarterly and even monthly counts may be possible in some large organizations and will allow more frequent decisions to adjust support, methods, resources, etc.

Figure 7. Counting and Timing to Measure Business Results.

formance. Figure 6 shows the average call-handling productivity of a group of service representatives trained in a fluency-based program accelerating to and beyond the level achieved by representatives who did not have fluency-based training.

Business Results

Figure 7 presents a few examples of measuring business results. Figure 8 shows the monthly revenues of an independent consulting business accelerating and becoming less variable over a period of 30 months—using the same charting system (Lindsley, 1999) as used in the previous figures illustrating measurement of learning and job performance.

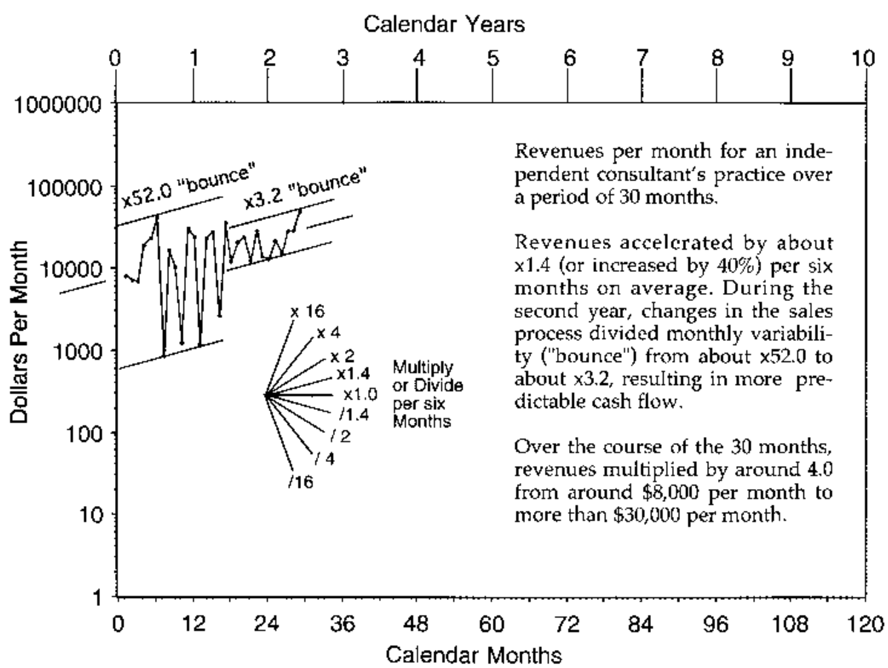

Figure 8. Monthly Revenues over 30 Months.

Conclusion

There is a fair amount of confusion about measurement in HPT—what it is and how to do it. Many programs and performance improvement efforts use such tools as rating scales, competency assessments, 360-degree surveys, and the like, combined with statistical analyses and graphic displays intended to accentuate (even exaggerate) the size of effects. Many of these approaches do not meet the criterion that a scientist, engineer, or accountant would use to design a measurement system: They do not use standard counting units for quantification and comparison of results. With so many approaches in the HPT marketplace described as "performance measurement," confusion is understandable about both the purpose and application of various types of tools and methods.

This article—by presenting and illustrating three simple principles for measuring and presenting performance data—is intended to simplify the job of the practitioner seeking to measure the effects of interventions. While it can still be *practically* difficult to identify countable units in many situations or to gather the data, my hope is that it will now be a bit clearer how to proceed. Find things you can count. Use standard units and standard time dimensions. Avoid presenting percentages or ratios without also presenting the original counts of behaviors, accomplishments, or results from which they are derived. Use multiply/divide factors rather than percentages to quantify differences and changes in performance.

Keeping to these principles will allow you to more easily evaluate effects, share data, and compare effectiveness across a broad range of interventions, improving both your practice and the field as a whole.

References

Binder, C. (1995). "Promoting HPT innovation: A return to our natural science roots." *Performance Improvement Quarterly*, *8*(2), 95–113.

Binder, C. (1996). "Behavioral fluency: Evolution of a new paradigm." *The Behavior Analyst*, *19*(2), 163–197.

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

Binder, C., & Watkins, C.L. (1990). "Precision teaching and direct instruction: Measurably superior instructional technology in schools." *Performance Improvement Quarterly*, *3*(4), 74–96.

Gilbert, T.F. (1996; Original publication, 1978). *Human competence: Engineering worthy performance*. Washington, DC: ISPI and HRD Press.

Johnston, J.M., & Pennypacker, H.S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lindsley, O.R. (1997). 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.

Lindsley, O.R. (1999). From training evaluation to performance tracking. In H.D. Stolovitch & E.J. Keeps (Eds.), *Handbook of human performance technology* (2nd ed.), (pp. 210–236). San Francisco: Jossey-Bass.

Skinner, B.F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Stevens, S.S. (1958). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: John Wiley & Sons.

Dr. **Carl Binder** is a partner in Binder Riha Associates, a performance consulting firm in Santa Rosa, California. He is best known for research in behavioral fluency and development of the FluencyBuilding™ training and coaching methodology and for contributions in knowledge management for sales and marketing organizations, performance measurement, and promotion of effective instructional methods for children. Having founded Precision Teaching and Management Systems, Inc., and Product Knowledge Systems, Inc. prior to Binder Riha Associates, he has consulted with scores of Fortune 1000 organizations, as well as educational and human services agencies. A long-time contributor to ISPI, he has authored several dozen articles and chapters in scientific, educational, professional, and business publications.

Carl had the great fortune to study with B.F. Skinner as a graduate student at Harvard and to learn from extraordinary mentors, including Beatrice Barrett, Ogden Lindsley, Eric Haughton, and Tom Gilbert. He attributes most of his good ideas to them, and to the measured results of his work with clients and learners over the past 27 years.

Carl lives with his business partner, Cynthia Riha, their three children, and a cat named Serena. His easy-to-remember email address is CarlBinder@aol.com, and you can download some of his previous publications at www.Binder-Riha.com.