

CLASSICS

# Measuring the Potential for Performance Improvement

By Thomas F. Gilbert

When we make judgments about the competence of human conduct, we often confuse behavior with performance. Behavior is a necessary and integral part of performance, but to equate the two is like confusing a sale with the seller. Naturally, we cannot have one without the other. But the sale is a unitary transaction, with properties all of its own; and we can know a great deal about it even though we know little—perhaps nothing at all—about the seller.

In performance, behavior is a means; its consequence is the end. And we seldom have any reason to try to modify other people's behavior in complete isolation of consequences. About the only reason would be to study it. By viewing behavior in convenient isolation we can learn many things about it, ranging from measures of visual acuity to useful information about the perseverance of habits. But those things *by themselves* tell us very little about performance.

## Measuring Human Competence

I believe wholeheartedly that any kind of performance can be measured—reliably and with considerable precision. We can measure the performance of poets, managers, teachers, physicians, lawyers, research scientists, psychotherapists, composers and politicians—not just that of production workers and athletes. The belief that the more complex forms of performance are not subject to measurement and quantification arises simply from ignorance about how to do it. Once you get the knack, per-

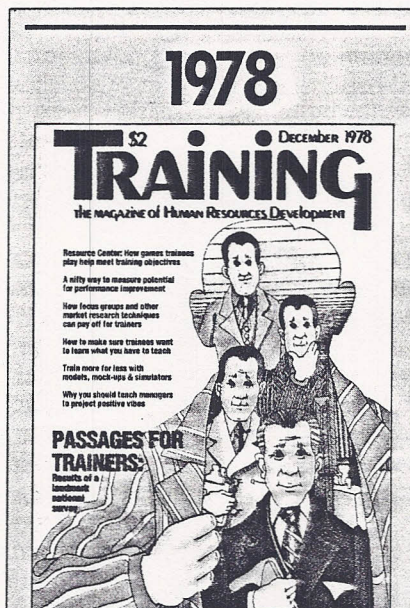
formance that you once thought unmeasurable will usually be not nearly so difficult to measure as, say, the radiation of Martian soil or the fertility of farmland. I hope to convince you

of this here.

But performance alone is not what I have set out to measure, because performance alone is not competence. Competence is a social concept, a comparative judgment about the worth of performance. In order to convert measures of performance into measures of competence, we require a social standard. Once we find that standard, competence will be as easy to measure as performance.

We get the competence of any one person, institution or culture only by comparing the very best instance of that performance with what is typical. Mark Spitz, the Olympic swimmer, was (at his best) only about 20 percent faster than the average high school swim-contest entrant, which means that the average high school entrant is exceptionally competent. Mark Spitz, of course, was a perfectly competent swimmer, because he was the exemplar. I call this measure of competence—the ratio of the exemplar's performance to typical performance—the PIP (potential for improving performance); and it doubly serves us. First, it tells us how much competence we already have; second, it tells us how much potential we have for improvement.

I define *exemplary* performance as the worth of the historically best instance of the performance. And notice that we need not accept mediocrity as a standard. For example, if a greenhorn's acre yields \$1,000 in grain at a cost of \$500, the typical worth index ( $W_t$ ) is two. If the best green thumb yields \$2,000 in value at a cost of \$250, the exemplary worth index ( $W_{ex}$ ) is eight. Then the green-



**If you were to list the people who have been most instrumental in the training profession's ongoing effort to transform its activities from an "art" into the "science" of performance engineering, Tom Gilbert's name likely would top your list. His 1978 book *Human Competence* is a classic. So is this article, adapted from the book for TRAINING in the same year (and slightly abridged here). As has always been his wont, he lobs some hand grenades at conventional notions of how to build an effective work force.—Editors**

horn's PIP is four, meaning that the greenhorn has the potential for doing four times as well. (Dollars are convenient units, but the PIP is by no means restricted to them.)

Human competence, then, is further defined by the Second Leisurely Theorem (or the Measurement Theorem), which states: *Typical competence is inversely proportional to the potential for improving performance (the PIP), which is the ratio of exemplary performance to typical performance. The ratio, to be meaningful, must be stated for an identifiable accomplishment, because there is no "general quality of competence." In shorthand, this theorem states that:*

$$PIP = \frac{W_{ex}}{W_t}$$

There is also an interesting corollary: *The lower the PIP of any person or group, the more competitive that person or group is.* Now, the word *competitive* is a delight to some people; to others, it signals unpleasant things. But that is because the cult of behavior has us confuse certain behavioral properties, such as greed, aggression, determination and the expenditure of energy, with competing. All I mean by competing is performing with comparative competence.

#### PIP Characteristics

You will note that the PIP is a measure of opportunity, the very stuff human capital is made of. The PIP does not assign feeble limitations to people as the IQ does but takes the humane and practical view that poor performers usually have great potential. Also, our measurement theorem does not posit competence mystically inside people's heads, but places it in performance. People are not competent; performance is. People have opportunities, which the PIP can express with precision and validity.

Indeed, the PIP can be measured as precisely and as accurately as we choose. Competence may vary from time to time, but our methods of measuring it need not. I have devised practical methods of measuring the PIP, and they need not be validated against criterion measures, because the PIP, when properly used, is the performance criterion. And, naturally, when applied in the world of work, the PIP yields accurate measures of the economic potential for

improving performance.

The PIP is principally a conceptual tool, which gives us a basis for comparing potential opportunities to improve performance. In general, the smaller the PIP, the less possibility there is to improve performance and the more difficult it is to reduce the PIP to 1.0. It is easier to reduce a PIP from, say, 4.0 to 1.5 than it is to shrink a PIP of 1.2 to 1.1. This rule is no longer true, however, if two circumstances hold. One is if we have full knowledge of why the exemplar is a superior performer, and we also have full control over those variables—that is, when we can give typical perform-

**You will note that  
the PIP is a measure  
of opportunity,  
the very stuff human  
capital is made of.**

ers the training, information, tools or motivation they require to emulate the exemplar. The second circumstance is even more important: when we can improve on the exemplar. Thus, the PIP is a "dynamic" measure, because new exemplary standards can always be set.

But here is something more to be considered. Even if we gave all performers the information, knowledge, tools and so on, of the exemplar, some variance in performance would remain—someone would still manage to shine as the best performer. In a "perfectly" competitive world, where we have arranged for everyone to have everything necessary in order to emulate the exemplar, such inherent characteristics as quickness, strength, "intelligence" and ambition will give some people a slight edge. In athletics, that slight edge is the critical distinction, but in the world of work or in the world of schools it would usually be of no special economical significance at all. It is, I believe, virtually impossible to reduce PIPs to 1.0, simply because someone will always discover a better way of doing it, have some natural superiority, or possess an unusual degree of motivation to excel.

What I am saying is that, in general, the more incompetent a person

or a group of people are, the easier it is to improve their performance. This contradicts the way we often think. But that is because we rarely think as performance engineers. Left to "nature"—to uncontrolled and unplanned events—exemplary performers are likely to improve themselves, setting new standards of exemplary performance. But as a situation becomes more "unmanaged," PIPs will grow—with the result that management has more potential for realizing them. Although large PIPs may discourage the uninitiated, they are a welcome opportunity to performance engineers.

The size of the PIP, of course, only indicates potential for improving performance—not how economically valuable that potential is. To put an economic value on a change in a PIP, we must translate it into what I call "stakes." (Stakes are the money value of realizing the PIP). A PIP of 4.0 in the speed at which janitors clean a building, say, does not translate into as much economic potential as a PIP of 1.5 in the speed of the production line. Later, I shall discuss the relatively simple techniques of translating PIPs into stakes; meanwhile, it is important to see the use of the PIP as a conceptual measure, pointing us in the direction of engineering opportunity.

#### A Case in Point

A case history, based on real events, illustrates how the use of the PIP can be a solid clue of great economic importance to a performance engineer faced with a really unfamiliar performance system. In this case, we shall see a performance engineer, Frank Roby, face an unfamiliar situation and find opportunities to improve it greatly—opportunities of the kind that experienced management misses every day.

The manufacturing vice president of Surfside Seasonings Inc., Willis Angel, is dissatisfied with the performance of his plants. He is determined to find some way to improve that performance, and he assigns three groups of people to conduct independent studies to tell him which programs he should invest in. Two of these groups are the corporate training and organizational development departments of Surfside. The third

group is a consulting firm specializing in management development. When Angel reads the three reports, he can hardly believe that the studies were independent because their recommendations are so similar. All three reports finger the first-line supervisors of the work force in the processing area as the culprits most responsible for the poor showing in the plants. The once stable, but aging, hourly workers have been largely replaced by young women from the ghetto. All three reports agree that the old first-line supervisors simply don't know how to manage the new breed. A training program in new styles of supervision, and in human relations, will be required; and the management consulting firm offers to develop one for \$78,000. For a \$400 million business, this does not seem too large an investment in good supervision.

Angel, of course, is impressed by the substantial agreement of the three studies he has commissioned. And the arguments have a certain face validity: The culture of the work force has changed, and there is no doubting that. But the \$78,000 training-development cost for an operation that has been losing money gives Angel trouble. He can't quite make up his mind, and he decides to get another opinion. He has heard of a consultant named Frank Roby, a man with a mixed reputation. Some say that Roby is completely without professional qualifications and imply that he is a charlatan. Others insist that although his methods are truly unorthodox, Roby gets results. The word *results* sounds sweet to Angel, so he hires Roby at \$750 a day.

Because of Roby's reputation, Angel decides to watch him work. Roby shows up one morning and makes the mandatory tour of a manufacturing plant, seemingly without noticing a thing. He then spends the rest of the day talking with the corporate accountant, the plant production manager and the chief quality-control inspector. To Angel's surprise, Roby appears in his office at 5:00 p.m. saying that he is ready to deliver his report and suggests that they conclude the study in the nearest bar over Vidalia Specials, a mixture of orange juice and sour mash bourbon.

While Angel begins his adaptation to this curious blend, he asks Roby if

he has ever been in a manufacturing plant like Surfside's. "Not exactly," Roby replies, "but I once helped some folks in a chewing gum factory."

So much for Roby's credentials. Angel begins the audition with deep suspicion, but after an hour Roby has completely convinced him that the best way for Surfside Seasonings to waste its time and money is to train first-line supervisors; and that, indeed, the company has an extraordinarily competent corps of foremen in the processing areas. (Mind you, Roby never so much as interviewed a supervisor.) Besides, Roby tells Angel exactly where he thinks the problem is, why it is there and what can be done about it. He is so convincing that the next morning Angel seeks authorization to spend the \$150,000 that Roby said would be required for the program.

Only 18 months later, Angel has sufficient data to prove that the adoption of Roby's program is netting the company a return of several million dollars a year in greatly increased labor productivity, decreased waste, lower employee turnover and fewer grievances. And Angel finds himself taking all the credit—not that he's that kind of guy. But how could he ever convince anyone that a man could walk into a seasonings plant for the first time and after a day tell you how to turn the plant around—and against all the advice of seasoned professionals?

We can look at just a sample of the data that Roby studied to reach his conclusions: Table 1 shows some pro-

duction data for three representative supervisory groups at Surfside Seasonings. (Of course, Roby didn't depend on these data alone, but they contributed far more than anything else to his remarkable conclusions.) In examining these data, Roby could see at once that the potential for improving the performance of the hourly employees was considerable, but that the differences among supervisors was small. Even though Supervisor B had the best supervisory performance in the company, getting other supervisors to perform as he does would not improve matters greatly. If the situation were reversed and there were large differences among the supervisors, his conclusions would have been quite different.

The average production is 96.93, and the best employee produces 194 units; so the employee PIP (assuming that costs and quality are the constants) is:

$$\text{Employee PIP} = \frac{194}{96.9} = 2.00$$

This employee PIP shows that the average hourly employee has the potential for doubling productivity. But the supervisory PIP is negligible—unusually low, in fact. Roby looked at these variances and then noticed that the job the employees had to do was to operate complex low-tolerance equipment. A lot of learning is required to master it. He also heard people say that it simply took a lot of experience to get maximum production. And he learned that the hourly

TABLE 1  
COMPARATIVE MANUFACTURING PRODUCTIVITY

Supervisor A		Supervisor B		Supervisor C	
Employee Number	Hourly Production	Employee Number	Hourly Production	Employee Number	Hourly Production
1	163	11	194	21	172
2	149	12	138	22	137
3	118	13	137	23	136
4	108	14	131	24	135
5	106	15	110	25	127
6	93	16	89	26	100
7	60	17	61	27	56
8	57	18	49	28	52
9	42	19	48	29	41
10	30	20	41	30	28
Average	92.6	Average	99.8	Average	98.4

employees got no formal training—mostly because production managers didn't think that formal training was as good as on-the-job experience. He considered this nonsense, of course, and he advised Angel that \$150,000 invested in proper training in the theory and troubleshooting of the equipment could get any new employee producing at about 150 units an hour, reducing the employee PIP to less than 1.3. Roby proved to be right—and the most important information he had was the PIP measures. Management had hidden the data in its books, but not in the form of Table 1.

Frank Roby is a real person, and this is an almost true story. It is true in every important respect except for the time it took—Roby has never met anyone as open-minded as Willis Angel; it usually takes weeks or months to build up sufficient appearance of credibility for his advice to be taken seriously. Roby has no magic, no mysterious capacity for insight. Indeed, his methods are so simple that when people watch his behavior, they cannot help but be unimpressed. Roby has learned to observe measures of competence and to make sense of them. Those simple measures can be powerful instruments in our pursuit of competence if we can set aside our behavioral biases long enough to see how they can be used.

#### Whose Performance Can Be Measured?

The Roby example deals with relatively simple performance that can be measured quite easily—in units of production. You might argue that much of the world of human performance is not so simple, and you might reasonably question whether other kinds of performance can be measured to yield neat units like the PIP. At least consider my proposition that any kind of performance can be measured.

Oh, the thrill when we first broke through what seemed to be the dense underbrush of John Donne's poetry. But if Donne was a competent poet, how can one measure that competence? Is there any way to say precisely that Donne is 2.3 times the poet that Herrick is, or 3.0 times Lowell? The obvious answer comes much too easily: There is no way to quantify beauty or spiritual power.

The problem is not *whether* we can measure Donne's performance, but

what it is we expect from poetry—what we consider a valuable poetic contribution. Measuring non-creative competence seems to be easier than measuring the competence of creative artists, but only because people can more easily come to an agreement about what is expected of non-creative performance. Anyone's performance can be measured in many different ways, and those measures become measures of competence whenever we can agree on what it is about the accomplishments that we value. . . . I, for one, must believe that it is possible to measure performance and competence—even John Donne's—and to make those measures mean something. It may not always be easy, but the stakes are reasonably high. Indeed, if we cannot measure competence, there is very little reason to talk about it at all.

The widespread feeling that many of the important characteristics of human conduct resist measurement is a result, I believe, of the familiar confusion between behavior and performance. There are at least two reasons why behavior is often difficult to measure satisfactorily: much of it is covert and not easily observed; and it is often hard to specify exactly what behavior is required for exemplary performance, because two exemplars may behave in considerably different ways.

If I were to build a scale of poetic competence, be assured that I would not start by observing the *behavior* of poets. Donne's poems speak for themselves—clearly to a few, not too clearly for most. No end of analysis of his behavior would add one scintilla to a proper assessment of his performance. Besides, as I have said, Donne's behavior is no longer available. And behavior is not competence any more than an eight-cylinder engine is a Sunday drive in the country. Once we lock that concept firmly in mind, it becomes much easier to measure human competence. T

Thomas F. Gilbert *now heads Performance Engineering, a consulting firm in Hampton, NJ.*

## THE MCGRAW-HILL SUPERVISION AND MANAGEMENT SERIES. YOUR COMMITMENT BEYOND TALENT, TO SKILL.

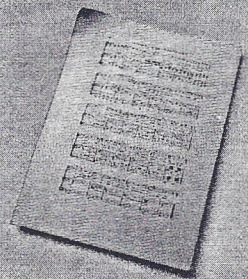
Working shoulder to shoulder with industries, corporations and government agencies across the country, McGraw-Hill Training Systems has developed the final word in supervisory and management training.

Two comprehensive series, available individually or together, as specific modules or a complete curriculum, off-the-shelf or customized to your specific business climate.

**Powerful. Intelligent. Effective.**  
Designed by pros to ensure success. If your goals are too high to be left to chance, your training partner is McGraw-Hill.

Call us today for the name and number of the representative in your area.

McGraw-Hill Training Systems  
1-800-243-5858



Write No. 190 on Reader Inquiry Card

**McGRAW-HILL**  
Training Systems

WORKING FOR GOALS. NOT LEAVING THEM TO CHANCE.