

**Assumption-Based
Metrics:**

*Recipe for
Success*

By David Apgar

Two challenges facing U.S. businesses raise the question of how well our performance management systems are performing. One: Although average worker productivity has risen briskly for two decades—and despite the growing power of enterprise planning systems, balanced scorecards, and quantitative analysis to measure success—average wages have stagnated. Unless businesses are systematically withholding compensation from workers, our performance systems are having a hard time pinpointing specific contributions to success.

Two: The flattening of U.S. consumer demand has put intense pressure on business leaders to pinpoint what drives success in new export markets. Yet early indications are that German and Japanese firms are adapting to new markets in places like China and India more quickly than U.S. firms.

It isn't for lack of data. Most U.S. businesses have too much of the stuff. Indeed, the main tool organizations use to focus attention is buckling under the pressure. In polls over the past decade, between 47% and 70% of executives expressed dissatisfaction with balanced scorecards—the performance dashboards that balance short- and long-term goals. The criticism is that balanced scorecards track too many irrelevant factors to provide focus. The challenge is finding metrics that matter.

Assumption-based metrics provide a more powerful way to determine what really drives success. Instead of collecting easily measured indicators in management reports that grow too voluminous for anyone to read, they begin with goal prioritization. The idea is to start with executives' intuitions about what currently matters most and then test those intuitions.

Assumption-based metrics represent a distillation of the best parts of literally hundreds of practices from some of the most effective companies and organizations around the world that my research teams reviewed over the decade that I launched and ran networks for finance executives at the Corporate Executive Board. The practice reviews were an effort to answer the question finance executives put to us most frequently: "How do I find predictive metrics?" It took years for an answer to emerge. This article summarizes a longer version of that answer in my book *Relevance: Hitting Your Goals by Knowing What Matters* (Jossey-Bass, 2008).

What Makes Metrics Predictive

Metrics need to have two things to help us predict results. They need to have a big impact on those results, and they need to have a fair amount of independence in the sense

that they don't depend on other things that have a big impact on results. For example, interest rates have a big impact on the demand for housing. This just means the demand for housing is highly sensitive to changes in interest rates. And if you are in construction, you'll probably want to consider interest rates in forecasting business.

Independence is a less familiar concept, but we use it intuitively all the time. For example, suppose you want to forecast your firm's share of the market for GPS systems in cars. You could measure customer satisfaction with your products and with those of your competitors. But it would probably be more useful to measure something like reliability because it drives both customer satisfaction and market share. In this case, reliability has more independence than customer satisfaction.

The search for metrics that have high independence and impact on the results we want to achieve is a search for root causes of success. High impact makes sure they are causes, and high independence makes sure they are root causes. It makes intuitive sense.

Yet most organizations abandon these intuitions when it comes to selecting and reporting on performance metrics. Since regressions are a good tool for testing the impact of indicators, we rely on business analysts to interpret the data that operating groups collect. Business analysts tend to have a great deal of computing power at their disposal these days, so it's easy to throw everything we measure into a regression and see what has high impact. In the process, we forget about independence.

The financial planning process reinforces this issue. Senior executives tend to work backwards from targeted results to what those targets require. The trouble is that requirements tell operating groups *what* to achieve but

Recipes pack a lot more information than requirements because they tell you how to achieve something. What recipes have that lists of ingredients lack is independence.

not *how* to achieve it. They are like lists of ingredients. What we need instead are good recipes. Recipes pack a lot more information than requirements because they tell you how to achieve something. What recipes have that lists of ingredients lack is independence. For example, hiring enough salespeople who can travel around the world is a requirement for an export business, but recruiting graduates from Indian and Chinese universities to sell into their home markets is a recipe. Cutting energy costs may be a requirement in a competitive market, but setting up a cogeneration plant is a recipe.

To get a sense of the difference between planning with recipes and planning with lists of ingredients, contrast some past episodes of performance management at Nestlé and Ingersoll Rand. When Herb Henkel first steered Ingersoll Rand away from its traditional base in locks, equipment, and electric vehicles to customer solutions, the firm's balanced scorecard reflected no single testable strategy or recipe for success. Instead, it looked like a list of broad requirements—necessary ingredients for various strategies but no coherent view of root causes of competitive advantage.

For instance, it aimed at both customer intimacy—critical for a solutions provider—and growth through innovation—a requirement for a product specialist. Led in part by those broadly conflicting metrics, the firm was soon acquiring specialist producers of parts that it could have procured for its clients at arm's length as a solutions provider. Analysts quipped that it provided great solutions at prices inflated by M&A premiums. Without metrics to reflect Henkel's specific vision, the firm drifted.

Under Peter Brabeck, Nestlé refined its recipes for success (no pun intended) every time it missed a target. The

company set broad cross-product goals for each of its regions and cross-region goals for each of its product groups. Managers then had leeway to shift goals across products or regions until goals for products within regions, and for regions within products, reflected equal tension. The result was detailed recipes for meeting regional goals across all products and product goals across all regions. Flaws showed up as conspicuous patterns of hits and misses across products and regions.

For example, suppose the KitKat and Nescafé global managers each expect to sell \$50 million of product in Vietnam and \$100 million in Indonesia. Knowing more about local tastes, the Vietnam manager might negotiate higher Nescafé and lower KitKat targets (say, \$70 million and \$30 million) if the Indonesia manager will accept lower Nescafé and higher KitKat targets (say, \$80 million and \$120 million). Now total country and product targets remain the same, but the specific product targets in each country reflect equal stretch or ambition given the market assumptions of the local managers.

As a consequence, missed KitKat targets in both countries are more likely to tell you something about the strategy for that product, and missed targets for both products in Indonesia are more likely to tell you something about the strategy for that country. Results relative to goals with equal stretch are more likely to provide clear lessons about how to revise the assumptions underlying their recipes for success.

Experimenting with Recipes

The search for root causes in the form of metrics with high independence and impact isn't easy. The challenge comes from independence. Highly independent metrics reflect highly specific ideas. For example, a measure of how many trust-based sales relationships you develop reflects a much more specific sales approach than a measure of how many sales calls you make. Because of their specificity, it's hard to be sure in advance whether these ideas will prove to be closely related to success or unrelated to it.

By contrast, it's often easy to tell from superficial resemblance whether a vague or general quantity is at least loosely related to success. You can guess that customer satisfaction is often related to profitability, for example, because the same things (whatever they may be) tend to lead to both. Concepts as broad as customer satisfaction, however, rarely give operating teams clear direction.

Whether you call them root causes, recipes for success, independent metrics, or tailored strategies, they need testing. There's no other way to find out whether specific

Making Strategy Experimentation Safe and Easy

Again, to find predictive metrics, you need to look for root causes of results. Yet root causes are highly specific. There's no way to tell whether a potential root cause is relevant without testing it. And testing through trial and error tends to involve lots of small failures before delivering big successes.

This may be the most promising role for technology in performance management. For example, a new mobile phone application called GoalScreen transforms strategy experimentation into a personal, fun, and private experience while incorporating the minimum rigor needed for users to find root causes in the form of metrics with high impact and independence. The idea is to learn from experience before, not after, the monthly performance review.

The GoalScreen process involves three steps: (1) identifying a business goal and your key assumptions about how to achieve it, (2) laying out metrics reflecting your assumptions, and (3) testing those assumptions. The process repeats as you sharpen your assumptions given new results.

The first step is to set a numerical target for what you want to accomplish over the coming week and list the key assumptions you make in thinking about how to do so. The fact that the GoalScreen process starts with assumptions makes it an assumption-based metrics system. As I mentioned earlier, in all versions of assumption-based metrics, some of your assumptions will be about things you can control and some about things you can't control. In the second step, you lay out metrics for each of your key assumptions.

Assumption-based metrics systems need to keep executives focused on issues that have high independence and a high impact on results. Accordingly, GoalScreen asks users whether the metric reflecting each assumption is independent. It also asks users for a best estimate of impact—and this makes the system unique.

For example, suppose you are trying to keep the cost of finished car dashboard assemblies to \$500 each. You believe this

cost is especially sensitive to two key drivers: the price of oil (because oil price drives the cost of the composites you use) and the defect rate on dashboards from your factory. For each driver, you need to provide an expected outcome, a worst-case outcome, and a worst-case impact (WCI) of that worst-case outcome on results. Let's say you enter expected outcomes, worst-case outcomes, and worst-case impacts for the two drivers as shown in Figure 1.

The values for expected and worst-case outcomes are self-explanatory. The +80 WCI for oil price means that if everything else goes according to plan but the price of oil rises to \$120 per barrel, then you expect the cost of one of your dashboards to rise from \$500 to \$580. Similarly, if all goes as expected except the factory defect rate rises to 4%, then you expect your dashboards to cost \$530.

This is all the information you need to run a performance analysis of your results for the week. Not only will it show you whether your expectations are right, but it will give you a sense of whether your assumptions about impacts are right and whether you may be missing a factor. Before seeing why this is so in the third step of the process, though, consider what your assumptions are telling you.

Your WCIs rank the impact of the issues you're tracking. After all, they are a direct measure of the worst damage to your results that a worst outcome for one of your assumptions can wreak. Since WCIs tend to fall off rapidly after the top four or five, you can quickly check

Figure 1

MY GOALSCREENS		
Oil price		
Expected outcome	100	\$/bbl
Worst-case outcome	120	\$/bbl
Worst-case impact	+80	\$
SAVE		

MY GOALSCREENS		
Defect rate		
Expected outcome	1	%
Worst-case outcome	4	%
Worst-case impact	+30	\$
SAVE		

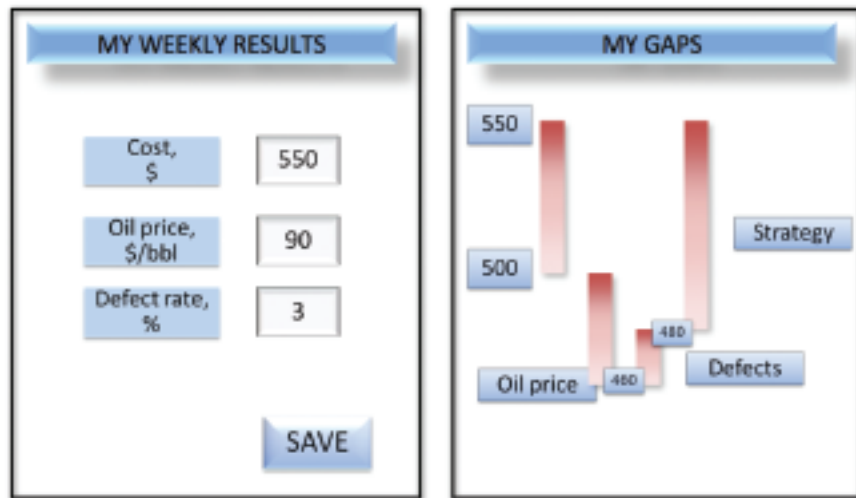
whether you are mixing unimportant ones with those that really matter by focusing just on the cluster of factors with the highest WCIs. In step three, you enter actual outcomes and results and get several screens of analysis. It's remarkable what you can learn from just a handful of numbers, but it's also the real point of the application. The application tells you everything you can learn about your assumptions from what just happened and sets you up to sharpen them for the next round. The following example continues from where the earlier one left off.

Costs were \$50 higher than planned even though oil prices were lower than expected. Clearly a defect rate not far from your worst case had a lot to do with it. The graph in Figure 2 shows the contribution of the outcomes for each metric to your result, assuming your WCI estimates are right (and also that the effects of deviations from your expected outcomes are simple proportions of them).

The oil price was better than expected by half the amount of your worst-case outcome, so GoalScreen assumes its impact should have been half of the \$80 WCI you estimated for it. In other words, the \$90 oil price should have lowered the cost of a dashboard to \$460. By the same reasoning, a 3% defect rate should have increased the cost of a dashboard by two thirds of the estimated \$30 defect-rate WCI—or from \$460 (given the oil price outcome) to \$480. Finally, GoalScreen gives the error—here the difference between the \$480 that your assumptions and the actual outcomes explain and the \$550 actual result—as a \$70 strategy gap. The strategy gap measures the part of the total performance gap that your assumptions can't explain.

Even in this simple example, you can draw three conclusions. First, you may have overestimated the impact of oil price because a better-than-expected outcome didn't keep costs down. Second, you may have underestimated the impact of defect rates because your actual result seems to reflect a poor outcome for that metric. Third,

Figure 2



you have a strategy gap. When a strategy gap is as large as the WCIs of the metrics reflecting your assumptions—here \$70 compared with an \$80 oil-price WCI and a \$30 defect-rate WCI—it's a strong indication that you are missing an assumption as important as the ones you've identified. Otherwise, many smaller factors would have had to turn out worse than expected at the same time, and that isn't as likely as missing a single important factor.

So GoalScreen shows you when you are probably missing an important assumption even though you haven't yet identified it. At this point, you need to review operations and see what might have gone wrong that could explain the surprisingly high cost for your dashboards.

In sum, your outcomes and results will show you whether your assumptions are justified. This is why it's safe to build assumptions into your performance metrics in the first place: The results correct them fairly quickly. Without those assumptions, you could never capture what is surprising about actual outcomes and results in enough detail to sharpen your understanding.

GoalScreen is only in the testing phase, but it's available free to IMA® members and readers through the end of the year. Just direct the browser of your computer to www.goalscreen.com. It also works for most cell phones. The site requires a quick registration to keep your interim results private. Please let me know any suggestions for smoothing it out and making it simple and enjoyable to use by e-mailing me at david@goalscreen.com. And keep testing your ideas.

Table 1 Goals and Expected Outcomes

Key assumptions (goals and risk factors)	Expected outcomes	Worst-case outcomes	Worst-case impacts
Baby carrier market share (ultimate goal)	10%	NA	NA
Defect rate	1 in 100,000	8 in 100,000	-4%

concepts matter to what you're trying to achieve. Organizations that make plans based on requirements and metrics that lack independence don't face this challenge. Their challenge is lack of direction. What, for example, could be wrong about making better sales calls? You can make a safe guess whether broad, vague, or generic metrics are related to success, but they won't tell people how to get things done.

The widespread reliance on generic strategies, plans based on requirements, and metrics that lack independence is one reason why you see little deliberate strategy experimentation anymore. After all, a company whose annual report promises to cut costs, explore market adjacencies, align supply chains, improve quality, and develop its people has little to test. But there's also another reason—one based on confusion and a little over-promising from the business analytics movement.

Business analytics refers to the application of data to business problems, which is a good use of technology. The confusion comes from the hope that models of results will become more predictive as you add more metrics, regardless of how predictive the individual metrics are. The problem is that there's no evidence for this. Although large sets of metrics can act like a metric with high independence, their collective impact deteriorates just as quickly as their independence increases. You can add the changing cost of paper, staples, and envelopes to your profit model, for example, but there's no reason to expect those new metrics to improve your ability to predict profit.

The temptation to use business analytics in place of tedious testing of new strategic ideas is overwhelming. Testing new ideas in the form of metrics takes time, may

lead to temporary wrong turns, and admits that we don't know everything. It's all a lot cleaner if the folks in the data lab could just print out a spreadsheet with the right answer. They can't, of course, but hope springs eternal. This is why scorecards of widely diverse firms look strangely similar.

Where deliberate testing of specific new strategic ideas and the independent metrics that reflect them does still occur, it tends to take the form of productivity programs. For example, Toyota's kaizen operating system asks work teams to identify the highest existing standards for their tasks, find ways to exceed them, revise the standards, and then repeat the process. GE's Six Sigma performance management system lets customers set those standards and then asks operating groups to cut all stages of the manufacturing process to the same low probability of missing them. In both cases, teams use trial and error to find new ways of accomplishing things and then codify their discoveries.

The IFC, a part of the World Bank that invests in for-profit companies, asks every investment team to devise new ways to get around risks identified in previous related projects at the start of each project. At the end of each project, the investment teams must write self-evaluations that include lessons future teams must consult. The U.S. Army employs a version of Rand Corporation's assumption-based planning in its counter-insurgency strategy involving the rapid testing and incorporation or rejection of new ways to enlist local populations. In both cases, organizations under pressure use trial and error rather than relying on data analysis because there's no better way to find specific strategies that work.

Assumption-Based Metrics

This is where assumption-based metrics come in. They make room to find the explicit strategy elements and risk factors that drive results in real organizations by harnessing business intuition and testing it—even if you don't have the luxury of randomized tests, control groups, and statistical methods. You can apply them at the divisional or corporate level of your company in parallel to, or even in place of, a financial planning system linked to balanced scorecards. They involve three logical steps.

The first step is to list every major assumption you must make to project a result against your ultimate business objective. Some of these assumptions will be about near-term goals that you control and some about risk factors that you can't control, but all are about drivers of success or failure. Along with each goal or risk factor is your expectation of an outcome. If your ultimate objective is to win 10% of the market for baby carriers, for example, then one assumption may be about the factory's defect rate, and your goal may be one defect in 100,000. List these goals and expected outcomes for risk factors in the first two columns of a table like Table 1.

The second step is to identify a worst-case scenario for each goal or risk factor. You'll need a different scenario for each one, but all the scenarios should be equally remote. In the case of defect rates, for example, the worst-case scenario may be failure of a key process control. Then estimate a worst-case *outcome* and a worst-case *impact* for the scenario that corresponds to each assumption.

The worst-case *outcome* reflects the impact of the scenario on your expectation for the goal or risk factor. Some people call this a *baseline*. In the case of factory

defects, for example, you may think the failure of a key process control could raise the rate to eight in 100,000.

The worst-case *impact* of each goal or risk factor is your best estimate of what the relevant worst-case scenario would do to your final results. Use judgment and experience—not new analysis unless it's readily available or widely known. In estimating the worst-case impact of each goal or risk factor, you should assume all of the other goals and risk factors turn out well.

In the case of an eight-in-100,000 defect rate, you may expect to win just 6% of the market rather than 10%. Write your estimates for the worst-case outcome and worst-case impact of each goal and risk factor in the third and fourth columns of the table.

Step three is the payoff. Simply rank your assumptions by worst-case impact. Most of the time, your impacts will drop off after the top four or five. Those are your priority goals, priority risk factors, and priority metrics—until the arrival of your next interim outcomes and results.

Those outcomes and results will show you more than just whether you're making progress toward your ultimate objective. They will also show whether your assumptions are justified. This is why it's safe to build assumptions into your performance metrics in the first place: The results will reveal whether your assumptions are good enough to stand for the time being or whether you need to change them. The sidebar on GoalScreen, a simple application of assumption-based metrics, provides a detailed example (see pp. 30 and 31).

Clarifying Success

Assumption-based metrics are just short lists of big-bet assumptions about what matters, but they let you extract critical lessons about what you are trying to accomplish. Moreover, assumption-based metrics can strip down your performance dashboard to a handful of key metrics and vastly reduce the time spent on financial planning and analysis. And they will relentlessly clarify what matters to success. **SF**

David Apgar is director of ApgarPartners LLC and a cofounder of GoalScreen LLC in Washington, D.C., and the author of Risk Intelligence (Harvard Business School Press, 2006) and Relevance: Hitting Your Goals by Knowing What Matters (Jossey-Bass, 2008). Before starting his own firm, he launched best-practices networks for CFOs, controllers, and treasurers as managing director for finance research at the Corporate Executive Board. You can reach David at (202) 337-1510 or david@goalscreen.com.

Further Reading

James A. Dewar, Carl H. Builder, William M. Hix, and Morlie Levin, *Assumption-Based Planning: A Planning Tool for Very Uncertain Times*, Rand, 1993.

Rita Gunther McGrath and Ian C. MacMillan, "Discovery Driven Planning," *Harvard Business Review*, July 1995, pp. 44-54.

Robert S. Kaplan and David P. Norton, "The Balanced Scorecard: Measures That Drive Performance," *Harvard Business Review*, January 1992, pp. 71-80.

David Apgar, *Relevance: Hitting Your Goals by Knowing What Matters*, Jossey-Bass, 2008.