

The Art and Science of Evaluating Organization Development Interventions

By Allan H. Church

How do we evaluate the impact of our organization change programs, processes, and initiatives? What are the best ways to measure success or failure of various interventions? How do we know we have really made a difference? While the field of organization development (OD) has its origins in action research and enhancing the growth and development of organizations and their people, if we are honest with ourselves, our focus as a field on formally evaluating the impact of our work has lagged far behind. The level of emphasis collectively placed as a field on the debate around having a clear and consistent definition of OD, the right set of core OD values, and creating new types of tools and techniques has far overshadowed the rigor and share of mind given to measuring the impact of our efforts.

Why is this the case? Is this due to practitioners' lack of measurement design and analytics capability as some have argued (e.g., Church & Dutta, 2013)? Recent research conducted on over 380 practitioners in the field (Shull, Church, & Burke, 2014) would suggest this might be part of the issue. Only 29% cited using statistics and research methods in their OD toolkit. Or is it because it is not part of their values structure? That same study reported that evidence-based practice was ranked 21st out of 34 possible values that drive their efforts. The history of evaluation as a core practice area in OD would also support this argument. Could the reason for our lack of focus on outcomes simply be because it's too difficult and daunting

to design a meaningful and robust evaluation process?

Whether you are an external consultant or internal practitioner there are a host of challenges associated with the measurement of causal relationships resulting from organizational interventions at the individual or systems level. Despite the definitive writings of Kirkpatrick (1998), the prospect of evaluating much of what we do in the social sciences and in organizational settings involves dynamic, interdependent, and often long-lead times that can far outlast the consultant-client (or employee-employer) relationship. In fact, one of the most frustrating elements of being an external practitioner is the lack of visibility to the long-term impact of one's work. This is often cited as a key reason why people take internal positions. While it's fantastic to experience multiple client organizations and challenges, it can also be quite rewarding to experience first-hand the changes in a company (or other social system) following work you have personally had a hand in over time. The alternate side of the equation are the personal challenges and even threat to evaluating one's own work. As an external consultant, if your project fails to deliver you might not get paid for that engagement. As an internal practitioner, your program or process might be cancelled and you could find yourself out of a job.

All of this sits squarely in juxtaposition with the client's interest in measuring the impact and in some cases financial return on investment (ROI) of our interventions. While this element of the work has always

been present (arguably few paid organizational interventions are done purely for the sake of humanistic values alone), in today's hypercompetitive business environment the emphasis has never been stronger. Given the dynamics and challenges cited above (e.g., capability, values, complexity, and personal interest in the game) how do we as OD practitioners move the needle and more holistically embrace the evaluation conundrum in our efforts?

The purpose of this paper is to discuss this issue in depth and attempt to answer these questions by focusing on the art and science involved in evaluating OD interventions. The paper begins with a brief overview of the evolution of the evaluation phase of the classic consulting paradigm from an afterthought to a core element required in the field today. This will be followed by a discussion of three key requirements for setting an evaluation strategy. Lack of attention to these areas works against practitioners and their clients' ability to appropriately conceptualize and implement suitable outcome measures. These requirements will be presented in the context of why they cause issues and how best to address them up-front in the process. The paper then offers three additional recommendations for creating an effective evaluation process that will yield the right kinds of information needed to demonstrate the impact of OD and related applied organizational interventions.

Recognizing Our Biases

Before proceeding, however, and as any OD practitioner should do, it is important to recognize several biases inherent in the approach to be presented here. First, the perspective is from that of an organizational psychologist by formal training with a significant grounding in applied statistics and measurement theory (and therefore represents a scientist-practitioner mindset). Thus, there is an inherent bias that evaluation methods for all types of organizational initiatives (whether directed at organization change, enhancing development, or improving performance) should follow some degree of rigor and contain

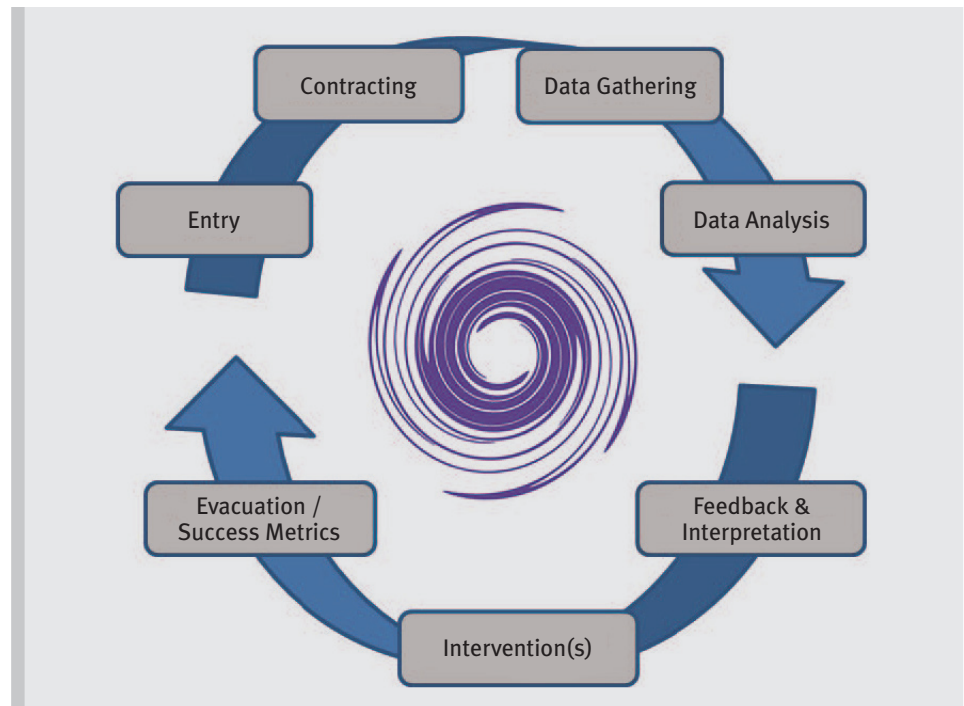


Figure 1. Classic OD Consulting Process Model

both quantitative as well as qualitative components of OD as a data-driven process (Waclawski & Church, 2002).

Second, the insights and observations offered are based on the author's personal external and internal consulting experience and evaluation research over the past 25 years following the implementation of a variety of large-scale organizational change initiatives and global talent management processes. The focus therefore is less on one-off individual coaching engagements, team building efforts, or group interventions but more on measuring the impact of broader initiatives, systems, and processes. Thus, while the suggestions offered reflect a certain normative and science-based paradigm, and are constrained by the experiences of the author, it is hoped that they will appeal to a much broader spectrum of OD applications.

The Evolution of Evaluation in OD

Although the evaluation stage of the classic consulting model that OD shares with many other applied social science disciplines, such as industrial-organizational psychology (I-O) and human resource development (HRD), has been present since Burke (1982), Nadler, (1977), and others (e.g., Rothwell, Sullivan, & McClean, 1995) discussed the framework, the

focus continues to be relatively limited in the field.

As many authors have noted (e.g., Anderson, 2012; Burke, 2011; Church, 2003; Cummings & Worley, 2015), the evaluation stage in the model is often given lip service or overlooked entirely. This is true whether you look at classic approaches to doing OD work as well as the newest dialogic approaches (Bushe & Marshak, 2015—where evaluation, for example, is not even listed in the topic index). A quick scan of the EBSCO database shows no academic articles published at all, for example, on the terms evaluation and OD since 2012.

Instead, the emphasis has often been placed on the actions taken or the change in behaviors and culture observed. While some research on the evaluation of various individual OD methods has certainly been done (e.g., Basu & Das, 2000; Terpstra, 1981; Woodman & Wayne, 1985), and there is a plethora of qualitative and quantitative case studies both individually and in OD texts citing the impact of various programs, relatively few authors have taken a more focused view of how to systematically measure our efforts. Moreover, many of these have come from authors with cross-disciplinary backgrounds (e.g., Armenakis, Bedeian, & Pond, 1983; Edwards, Scott, & Raju, 2003; Martineau & Preskill, 2002;

Rothwell, et. al., 1995). At the same time, scholar-practitioners in other related disciplines (e.g., Holton, 1996; Kirkpatrick, 1998; Roberts & Robertson, 1993; Svyantek, O'Connell, & Baumgardner, 1992; Terpstra, 1981, Woodman & Wayne, 1985) have taken this topic on years ago. Surprisingly enough of those with more academic orientations have advanced the field in far more significant ways than more traditional OD scholar-practitioners with the introduction of scorecards (e.g., Becker, Huselid, & Ulrich, 2001) and more recently bottom-line linkages (e.g., Lawler, 2014;

Many practitioners have as their goal culture change, behavior change, process improvement, organizational effectiveness, enhanced team functioning, etc. These are all excellent objectives but they are not tight enough to be used as measures of impact. You need to be able to answer the question of “what will be the measurable indicators of a positive outcome as a result of this effort.” These can be quantitative or qualitative (some of both methods are generally best for maximizing perceived evaluation credibility) but they need to be measurable and aligned.

Savitz & Weber, 2013), and the application of decision-science (Boudreau & Ramstad, 2007) to their work.

Given the trends in the field it would seem we are at cross-roads. There is increasing pressure to measure the impact of our efforts (Anderson, 2012; Shull et al., 2014), at multiple levels in the organization (Lawler, 2014), and using more types of complex and potentially anti-OD oriented “Big Data” applications (Church & Dutta, 2014). Yet arguably few OD practitioners have the right set of capabilities included in their formal training (Church, 2001) or core toolkits (Church & Dutta, 2013; Shull et al., 2014) to address these trends. Consequently, we are simply not engaging in rigorous evaluation methods of our organizational interventions. Let us now explore the reasons in more depth in the hopes of finding some answers and potential solutions to this challenge in the field.

Three Key Requirements for Setting an Evaluation Strategy

There are many reasons why OD professionals might not pursue or even actively seek to avoid engaging in the evaluation stage of their work. Some of these reflect internal states and motivations (e.g., values, personal investment in the outcome) and/or a lack of specific capabilities and skills (e.g., in research design and statistics). While very important for setting a baseline, they do not necessarily tell the complete story. Instead, the

emphasis here is on a third set of reasons—that is, the dynamics and complexities of measuring change real time in organizations. Listed below are three key challenges and requirements for setting an effective evaluation strategy.

1. Clarifying the Definition of Impact

Whether we start with a measurement theory approach (what is the criterion?) or an OD consulting model (what is included in the contract?), the importance of clarifying up front the outcomes to be measured is critical to the success of any intervention (see *Figure 1*). This is true whether the effort is a simple team building exercise, an executive coaching assignment, the implementation of a large-scale engagement survey program, or a whole systems process intervention. Whatever the initiative, it is imperative that outcomes be clearly articulated and there is alignment

up front in the contracting or project charter stages. Importantly, this is more than a set of objectives or goals for the project. The definition of impact, i.e. the intended outcomes, needs to be crystal clear in such a way that it can be measurable at one or more specific levels of analysis.

Many practitioners have as their goal culture change, behavior change, process improvement, organizational effectiveness, enhanced team functioning, etc. These are all excellent objectives but they are not tight enough to be used as measures of impact. You need to be able to answer the question of “what will be the measurable indicators of a positive outcome as a result of this effort.” These can be quantitative or qualitative (some of both methods are generally best for maximizing perceived evaluation credibility) but they need to be measurable and aligned.

Marshall Goldsmith, for example, in his coaching practice is known for his contracting efforts around behavior change. As part of the commitment to his work, he uses a pre-and post-behavioral feedback assessment tool following a yearlong engagement. If he does not see change in the measure, it is a direct indicator of the success or failure of the coaching project. While this might sound simple enough there are important measurement aspects, such as the quality of the feedback measure used, the nature of the raters selected, the rating process (confidential vs. anonymous) etc. that can impact the outcomes in ways that might be unexpected.

While other interventions (e.g., a cultural integration effort following a merger and acquisition) are much more complex than this, the same principle applies. In this context, the outcome measure might be comprised of targeted levels of turnover, identification and retention of key talent, improvements in levels of engagement or other cultural indicators on an employee survey, increased performance in business units, or an increase in the innovation pipeline 2–3 years later. There are no correct answers but there are aligned ones. The focus needs to be on the realistic measurable indicators of impact

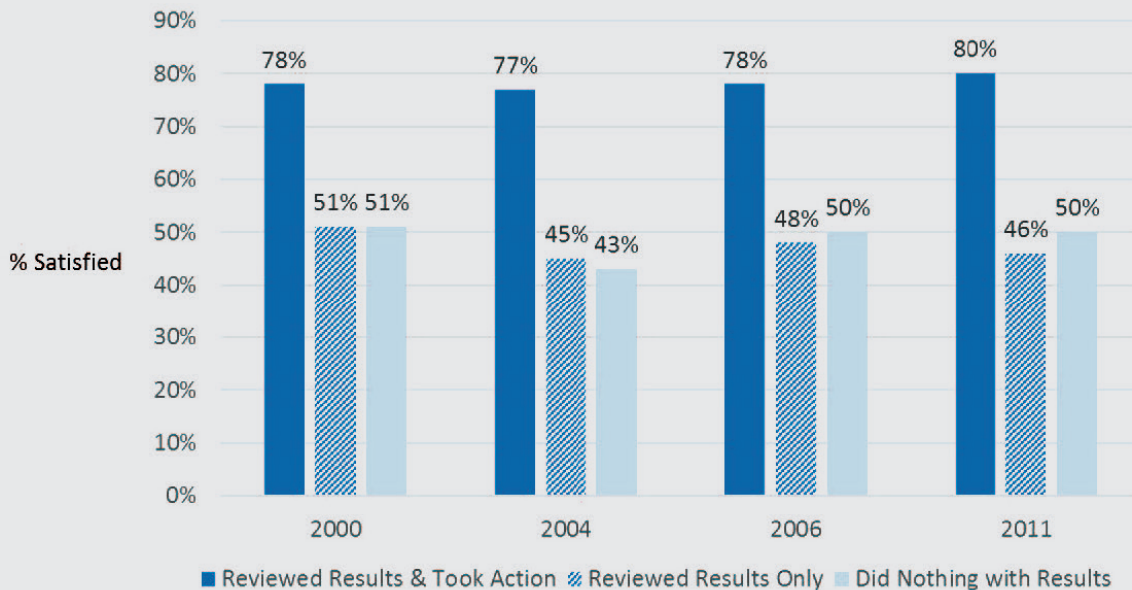


Figure 2. The Impact of Taking Action From Survey Results on Employee Satisfaction

that can be linked to the timing of a specific intervention.

2. Setting Realistic Time Horizons for Measurement

OD practitioners all know the simple fact that change takes time. Based on the prevailing problem, scope, and intervention, this can range anywhere from minutes following a process observation to years after a leadership transition. Unfortunately, clients are not always of the same mind. While fewer and fewer executives seem to believe in the fallacy of changing corporate culture overnight, their sense of timing and urgency is often directly proportional to the pace of their business. For example, consumer products organizations generally move faster than pharmaceuticals. The point is that as part of the outcome alignment process OD professionals need to ensure that the timing window of the evaluation and measurement components is clear and reasonable.

Another issue that can occur is an over emphasis up front on planning for the timing of the intervention launch, and less attention paid to the appropriate lag time required to observe the impact of the effort. Often this is because the bulk of the development work and consulting delivery costs are front-loaded. There is more concern about meeting the deadline to deliver a new program or rollout a change agenda

on target than there is on determining the best window (and method) for measuring the impact of that work over time.

A related issue, and common fallacy in organizations, is the use of the “pilot” concept as a means for testing the impact of a new program. While launching an intervention in a small-scale environment or controlled area of the business can be very useful for ironing out the implementation kinks, rarely does this offer an effective means for predicting the potential impact of a much larger scale program. This is because larger scale OD efforts need to be aligned to a larger set of systems factors which require much broader thinking about organizational impact than what typically occurs in a small pilot context. The question to ask yourself here is “given what we are anticipating measuring, when do we expect to see this outcome change as a result of our efforts?”

One important caveat should be raised here. Although the discussion so far might suggest that all OD interventions have a distinct beginning and end to them, we know this is not the case. While the classic consulting model tends to present the world in this semi-linear fashion, the vast majority of our work rarely begins from a blank slate. Burke, et al. (1997) have termed this effect “midstream consulting” and it applies in just about every case whether internal or external. Further,

there is often not a definitive end to the engagement. In fact, many organizational processes (and in particular those implemented for employee development, performance management, and talent management purposes) continue to evolve long after the initial design and implementation phases. From an evaluation perspective then the measurement aspect of assessing impact needs to be seen as occurring at discrete points-in-time and not as an end-state. This is an important distinction as it enables the practitioner to contract regarding “points of impact” measurement at different stages of evolution, and not rely exclusively on a single evaluation metric. Not only should this remove some of the burden of having to show impact all at once, but the measurement quality will improve as well. Time series studies and multi-method approaches are far more rigorous and valid than are single program reviews.

Case in point, in the mid-2000s there was an applied study done at PepsiCo on the impact of their global employee organizational health survey program (Church & Oliver, 2006). The research was conducted in an effort to answer senior leadership’s questions regarding the impact of the survey on key employee outcomes. The researchers analyzed survey data over time, including the use of an action planning variable, and demonstrated the impact

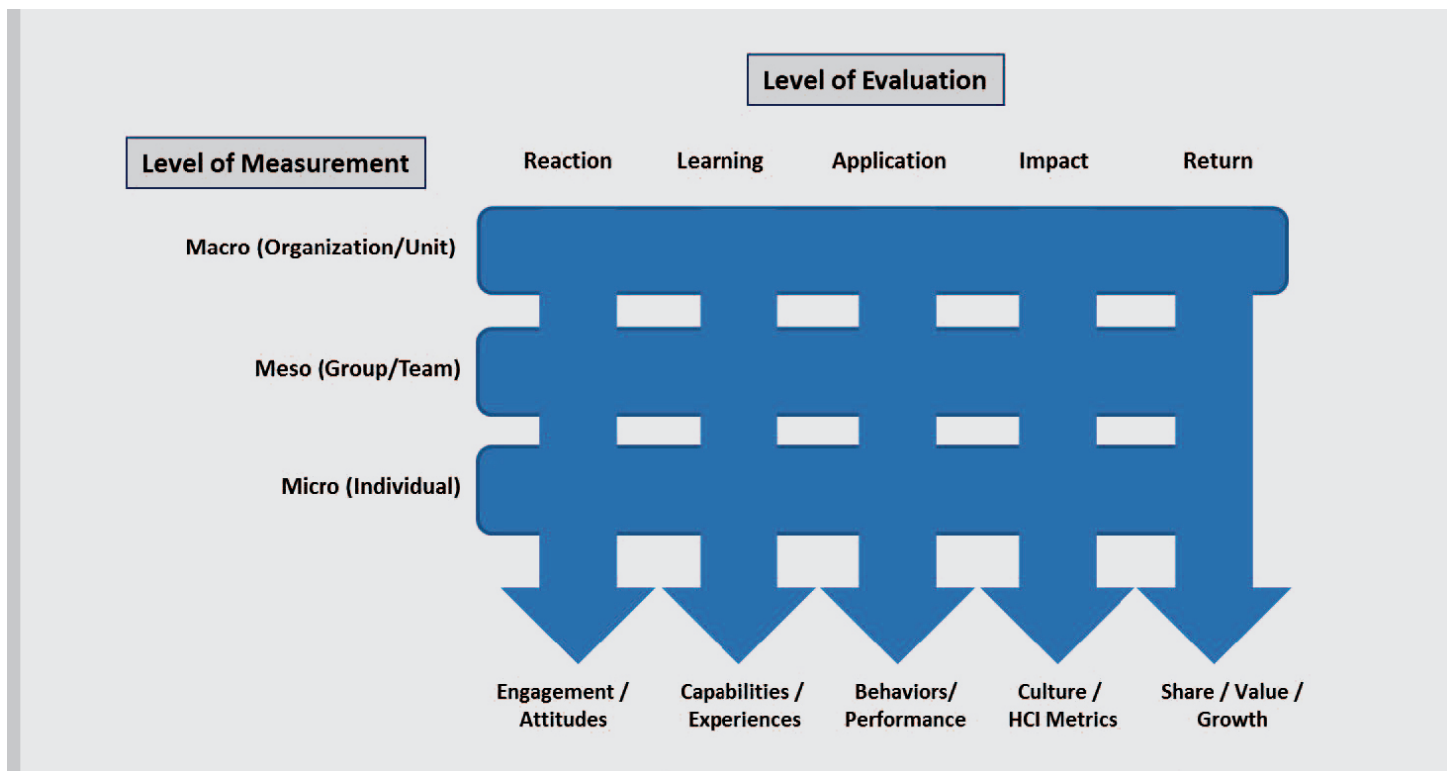


Figure 3. A Multi-Level Framework for Aligning Evaluation Efforts

of taking action from the results on both softer survey outcomes of employee satisfaction and commitment as well as hard metrics such as turnover, lost time and accidents at the plant level. Specifically, they reported that employee satisfaction and commitment a year later (via the second survey) were significantly impacted by managers who both shared the results and took action versus those who only shared results or did nothing at all. While these insights were extremely well received in the organization, several years later the same questions emerged under new leadership, so the study was conducted again (Church et. al., 2012). The same results were evident across multiple years (see Figure 2) demonstrating the power of an organizational survey with an action planning focus at the local level to drive organizational change and employee engagement.

3. Applying Systems Thinking to Systems Interventions

This third recommendation is simple enough. OD practitioners have a deep understanding of systems thinking so it should be easy to apply that same approach to the evaluation of their efforts. This means considering variables across all levels and sub-systems involved (Katz & Kahn, 1978) and aligning the measurement

of those impacting and impacted by the change. Whether you prefer the Burke-Litwin Model (1992) or some other approach to conceptualizing an organizational system, it is critical that broader thinking be applied than just a micro analysis of a single intervention. This is actually one area where OD practitioners should have the advantage over practitioners from other disciplines such as I-O Psychologists who tend toward a more individual level of analysis. Unfortunately, OD practitioners seem to ignore their own strengths when it comes to evaluation.

Too many practitioners treat the evaluation process as an afterthought or something to be considered once they are further along in the intervention. Though there is certainly wisdom in revising, aligning, and adjusting the measurement approach if needed as the intervention progresses, it should have a solid basis in evidence-based science articulated at the beginning of the intervention. When practitioners fail to focus on evaluation early in the process, it is no surprise that when pushed by clients to show results, they need to scramble to put something in place. This often results in a poorly designed and/or implemented measurement approach which can yield inaccurate results and might even derail the intervention itself.

As it turns out, often the best form of evaluating impact is to design a new measurement process at the start of the intervention or change effort to enable a pre-post comparison. The Marshall Goldsmith approach is in fact a perfect example of this principle in action as is the PepsiCo organizational health survey where the questions regarding action planning were asked the year before the company was interested in learning about the impact of the results. Even if you are not in a position to implement a pre-measurement tool, there are scenarios where it might take significant time and resources to collect the necessary information to show results, and this would need to start earlier on in the effort. Take for example, an organization that would like to know the impact of a new leadership curriculum on online learning utilization. While the data regarding learning system utilization might not be centralized at the start of the project, by the end centralization would be necessary to aggregate the data. This, would require lead time and dedicated resources.

Kirkpatrick's famous multi-level framework (1998) is perhaps the best and most easily recognized approach to setting a systemic strategy for evaluation. While initially designed for the evaluation of learning interventions it has since

been expanded to include additional concepts such as ROI (Phillips, 1991) and can easily be adapted to OD and related work including talent management (e.g., Church, 2013; DeTuncq & Schmidt, 2013; Holton, 1996). The core idea is that there are multiple levels of impact that can be measured in various ways (as aligned and timed per above). Essentially the model measures outcomes at the following levels: (1) reaction, (2) learning, (3) application, (4) business impact, and (5) bottom-line/long-term outcomes.

The key is spending the time required at the initial stages of the effort to design an integrated systems approach to the five levels of analysis, with the right types of metrics and under the right time horizons. This needs to be measurable to: (a) have sufficient rigor to demonstrate the impact of your efforts, (b) satisfy your clients, and (c) be reasonably executed with enough latitude to adjust for potential contingencies. *Figure 3* provides a simple framework for how this might be applied.

Only by setting the strategy for evaluation up front can you get ahead of the potential issues that will naturally come with these types of efforts.

Three Recommendations for Building an OD Evaluation Process

Now that the importance and critical factors involved in having an evaluation strategy have been discussed let us turn to the process itself. What elements and types of data should be included in the evaluation process? What are the key factors in conveying the messages around the impact of your efforts? What is the role of OD values in analytics (is that an oxymoron?) and what are the pitfalls to avoid? Listed below are three key recommendations for developing an impactful evaluation process. While they are not meant to reflect all the elements of what's involved in designing evaluation research (see instead texts such as Edwards et al., 2003; Kirkpatrick, 1998) these three should be helpful in designing an evaluation approach that helps put your OD efforts in the best possible light yet stays grounded in solid measurement theory and rigor.

1. Design Using a Multiple Measures and Levels (MML) Approach

One of the aspects that makes OD efforts so exciting and attractive to people as a profession is the variety of projects that comprise the spectrum of our work. Whether the interventions are focused on the individual, group, or organizational levels, because they are grounded in the social and behavioral sciences there are always a myriad of complex human dynamics involved (unlike say pure management consulting which often focuses more on business strategy, design, or financials).

By using multiple levels of analysis, you enable a more complex and interdependent way of assessing impact and change. So, for example, from a multi-levels approach, in measuring the rollout of a new set of corporate values you might measure both how senior executives model the behavior in town halls and other public forums (via personal observation or interview feedback), how middle managers are rated as demonstrating the behaviors in the workplace (via 360 feedback), and how employees feel about the authenticity of the new values for the culture (via employee focus groups or a pulse survey).

Unfortunately, this element of OD work also makes it particularly challenging to evaluate at times. Often the interventions are focused on less quantifiable aspects of human interactions such as group dynamics, power, archetypes, norms, and culture. Even when behaviors are involved (e.g., such as leadership competencies, management practices, communication or collaboration skills, digital capability, learning, etc.) they are not always easily measured by standard tools or off-the-shelf assessments. In addition, the measurement of performance has come under fire recently in the literature (e.g., Church, Ginther, Levine, & Rotolo, 2015) as being negatively received by leaders and poorly designed and implemented in companies today. It is no wonder then that recent conversations with senior leaders have resulted in their questioning the use of internal performance management data as a valid indicator of the impact of OD interventions. If PMP is not measuring the right things how

can it serve as a criterion for something else? The bottom line here is that to have an effective evaluation process in today's multi-faceted data-driven landscape you need to have a multiple measures and levels (MML) solution.

What does this mean in practice? Maximizing your ability to measure the true impact of your efforts requires more than one type of tool, process, or information system producing data. Preferably these multiple measures are done at different levels of analysis (e.g., organization-macro, group-meso, and individual-micro), and

you collect at least two different types of data at any given moment in time. By using multiple levels of analysis, you enable a more complex and interdependent way of assessing impact and change. So, for example, from a multi-levels approach, in measuring the rollout of a new set of corporate values you might measure how senior executives model the behavior in town halls and other public forums (via personal observation or interview feedback), how middle managers are rated as demonstrating the behaviors in the workplace (via 360 feedback), and how employees feel about the authenticity of the new values for the culture (via employee focus groups or a pulse survey).

Conversely by using a multiple measures approach at the same level and point in time you enable a process of triangulation. This allows you to see if all measures are showing the same type of change as a result of your intervention or if one indicator is moving when another is not (or

going in the opposite direction). Continuing with the values rollout example, from a multiple measures perspective in order to add a second indicator of the practice of the new values at the managerial level, you might incorporate an audit of performance reviews along with the 360-feedback process. The question would be are managers being reviewed by their bosses based on the values, and are they being rated by others as demonstrating them? A nice side benefit of this approach would be the correlations you could run between manager ratings on the 360 survey and manager ratings (if available) on the performance tool. By using this MML approach applied to OD interventions you are essentially following a similar path to what I-O psychologists call the multi-trait multi-method (MTMM) which is used for assessing individual skills and capabilities.

Survey feedback, behavioral feedback, performance ratings, observations, interviews, and focus groups, etc. are just some of the ways you can collect data and use them for your evaluation purposes. The practice of OD is replete with all sorts both qualitative and quantitative data-driven tools (Waclawski & Church, 2002) and the savvy applied behavioral scientist can convert the output of almost anything into a measure that can be used for some level of analysis (even if it's non-parametric). Linking that softer data to perceived harder metrics such as the those in the list below is where the rubber meets the road:

- » performance (caveats notwithstanding)
- » turnover
- » quality of new hires
- » promotion rates
- » talent transfers
- » external reputation indices and awards
- » market share
- » EBITA (Earnings before interest, taxes, and amortization)
- » diversity representation
- » customer satisfaction data
- » learning completion rates
- » sales
- » productivity
- » product shrinkage
- » senior leadership tenure, etc.

For example, in evaluating the impact of PepsiCo's *Potential Leader Development Center* (PLDC), a custom talent assessment process and part of their broader LeAD program aimed at identifying high potentials early in their careers, the program team employed a MML approach (Church & Rotolo, 2016). Specifically, they wanted to know whether being transparent (i.e., sharing the results of the assessment of "potential") with over 5,000 employees globally had any impact on how people felt about the program itself and/or resulted in any unintended changes in turnover or performance. As part of the evaluation strategy, at six months and one year after the individual feedback reports had been delivered, the team surveyed participants regarding their attitudes about the program content and mechanics and their perceptions of the feedback they had received. The data was then linked at the individual level to assessment results (i.e. leadership potential scores), annual business performance ratings, promotion rates, and turnover. After an in-depth analysis, the project team was able to answer senior leaders' key questions regarding the impact of the program one year later on talent movement, the extent to which results were used in individual development planning, and on employee perceptions of the culture.

More specifically, the research indicated that: (1) the assessment process was effective at predicting future success—i.e. actual performance and promotion rates one year later were significantly correlated with performance on the assessment tools; (2) transparency of how employees scored (their level of LIFT, a proxy for potential) had no negative impact on satisfaction with the program (70% favorable), perceptions of organizational commitment, or actual turnover; and (3) the program met its goal of providing developmental feedback to all participants with the vast majority (77% and 83% respectively) indicating that the results had helped them increase their effectiveness as a leader and showed an investment by the company in their personal growth and development. In short, the data provided statistically significant and meaningful results regarding the impact of the program and

dispelled management's myths regarding the potential negative outcomes of being transparent with the results. It would have been impossible to demonstrate these relationships without using this type of MML evaluation approach.

In the end, there is no single best outcome measure or set of measures. The important point is to always cycle back to your evaluation strategy and measurement construct and build from there laying the foundation for multiple measures. It is also important to remember, however, two old adages when it comes to measurement: (a) what gets measured gets done, and (b) beware the law of unintended consequences. If the goal of your talent management program is enhanced movement and you do not account for other factors then you will get movement even if it is of the wrong type (Church, 2013; Church & Waclawski, 2010). Remember to think holistically and at the systems level when designing your measurement process.

2. Build a Meaningful Story Through Insights

In as much as the first recommendation is about getting your hands on a significant amount of data from different sources, the second key to creating a sustainable evaluation process is the ability to build a meaningful story out of the insights you generate from that data. There are two parts to this.

First, having data by itself with no connectivity or insights is meaningless. It's just information. This is true whether it's one simple piece of evaluating ratings or five years of culture data linked to business unit performance and employee turnover. How significant is the overall effect? Where is the change really happening and where is it failing? What are the dynamics and interplay between culture, behavior, and turnover? All of this needs to be answered in a way that answers the questions posed at the start of the intervention.

The data you gather for your evaluation efforts must be analyzed in such a way to create insights into what is going on in the organization. In terms of the values rollout example, if the 360-feedback data were to show that managers were engaging

in the desired behaviors but the reward systems were not reinforcing them, this would be a key insight and useful in explaining why the program could be failing or having less of an impact than might be desired. This is why analysis skills though important by themselves are not enough (Church & Dutta, 2013). The ability to see connections and derive meaning from those connections, in some cases even causality between variables if the methodology and research design support these conclusions, is critical. It is a skill that comes with practice and experience working with large and often complex datasets (Church & Waclawski, 2001).

In the end, there is no single best outcome measure or set of measures. The important point is to always cycle back to your evaluation strategy and measurement construct and build from there laying the foundation for multiple measures. It is also important to remember, however, two old adages when it comes to measurement: (a) what gets measured gets done, and (b) beware the law of unintended consequences. If the goal of your talent management program is enhanced movement and you don't account for other factors then you'll get movement even if it's of the wrong type . . .

Second is the ability to tell a compelling story from those insights. Here your communications skills are tested. Even if you have the greatest dataset in the world regarding your OD intervention, if you put it in front of senior executives they may not know what to do with it. In fact, the more complex your data the worse it gets. A collection of interesting insights alone is often not enough to determine your intervention's success. You need to be able to put it all together into a compelling package that is tailored to the right audience (Church & Waclawski, 2001). What has been happening since you launched your intervention? What other changes at the systems level have occurred? How have these affected the change program and how do you know that (e.g., what other measures or data do you have)? Can you pinpoint the drivers

of impact and identify clear areas for action going forward?

Although the approach outlined here is similar to the data analysis, feedback, and interpretation stages (aka diagnosis) of the classic OD consulting model, there is a slightly different spin. While the same analytical techniques and capabilities generally apply, the emphasis in evaluation is on what has changed both positively and/or negatively as a consequence of your interventions. The point is to isolate the direct and indirect results of your efforts, and the potential moderating effect of other factors that you have measured over time as well.

3. Maintain a Watchful OD Values Lens to Your Evaluation Work

The final recommendation for building an effective evaluation process concerns a return to the use of OD values. This paper has not been focused on the OD values component of evaluation, yet it is vital to ensuring our work is evaluated effectively and in the right context. While there has always been an element of the dark side involved in data analysis and storytelling (see *How to Lie with Statistics* by Huff, 1993), the rise of Big Data and analytics functions in organizations is exacerbating this problem exponentially. There are several reasons for this situation.

First, the fascination with Big Data and machine intelligence is making all types of analyses even more complex. OD interventions and I-O practices such as selection and high-potential identification are now

part of the ongoing debates in the field (e.g., Chamorro-Premuzic, et al., 2016; Church, 2014; Church & Silzer, 2016; Rotolo & Church, 2015). As information flows up and down at an increasingly rapid pace (remember Big Data is characterized by volume, velocity, variety, and veracity) the challenge of determining causality versus random relationships is even more apparent. Unfortunately, many analytics teams take the data blender approach (e.g., throw a large cluster of variables in a blender and mix it all up and see what spits out) and the resulting findings can be meaningless or at best difficult to interpret. This suggests not evidence-based science but more of a fishing expedition. Relationships that make little sense can emerge and limited judgement is applied. Without a values filter applied to the analysis of data, relationships identified might be spurious, or even potentially unethical from an OD perspective.

Consider the case of employee engagement survey data that was collected under the auspices of being confidential, yet through linkage research is connected at the individual level to other variables such as performance data, promotions, social media postings, and iPhone activity. If these analyses are done internally and misused (e.g., to classify individuals for various decisions or opportunities) it could result in a major violation of employee trust and engagement with the organization. Once it gets around that leaders cannot be trusted, the entire process of gathering employee perspectives will be destroyed. This is just a simple example but a powerful and quite real one. As OD practitioners, we need to ensure that the data is used in the way it was intended and communicated to employees. The psychological contract and integrity of our profession needs to be maintained. One misstep with employees can erase in a heartbeat the positive momentum gained from an entire multi-year OD intervention.

Big Data by itself of course is not the problem, nor is the use of advanced analytics capabilities. In fact, the future of human capital management is going to require a digital mindset and statistical prowess beyond what most practitioners

have in their toolkits today. Rather the problem rests with those doing the analyses. The concern is the application of what might be termed “values free analytics” to understanding the social organizational phenomena at work. While linkage and similar research represent important approaches for demonstrating impact, in the wrong hands they can be misleading or even damaging to an organizational change agenda and threaten the existing culture and practices.

The key recommendation here is to ensure that whatever analysis approach is taken safeguards are in place to protect employees and the organization within the context of the work being done. OD practitioners should play an integral role in the analyses being done (if not doing them ourselves) to ensure they are done in the right manner and with the right frame of reference. This means being involved in every phase along the way from designing the evaluation approach, selecting the outcomes, asking the right questions, ensuring the analyses are robust and appropriate to the data in hand, interpreting results to determine key insights, telling the story given the context involved, and working with the client to make the right decisions. None of these steps should be left to someone with a limited sense of OD or complex change dynamics (e.g., a pure statistics or economics background). We bring a unique perspective to the table as OD practitioners with a specific set of research questions, and we should always be present to influence and protect the use and misuse of employee data.

Conclusion

The purpose of this paper has been to discuss some of the challenges and opportunities associated with the evaluation of OD interventions in organizations. Historically, the evaluation component of the classic consulting model has largely been downplayed or ignored. Today, with increasing pressure from organizations to demonstrate the value of our efforts, having both a well-designed and articulated evaluation strategy is key, as is a detailed multi-measure and multi-level measurement

process. While many OD practitioners continue to have limited interest and/or capability in engaging in evidence-based science for their evaluation efforts, with the increase in Big Data and the need for ROI there is nowhere to hide anymore. Going forward practitioners need to start with the science to build these skills and ensure they are designing the right types of measures and analyzing the data appropriately. OD professionals need to be facile at generating both insights from their data and the ability to tell a compelling story about the impact of their interventions. There is clearly a science and an art to conducting evaluation efforts in OD, and none of it should be as painful or as daunting as one might think.

References

- Anderson, D. L. (2012). *Organization development: The process of leading organizational change* (2nd ed.). Thousand Oaks, CA: Sage.
- Armenakis, A. A., Bedeian, A. G., & Pond, S. B. (1983). Research issues in OD evaluation: Past, present, and future. *The Academy of Management Review*, 8(2), 320–328.
- Basu, K., & Das, P. (2000). Evaluations dilemmas in OD interventions: Mixed record involving Indian rural credit institutions. *Public Administration Quarterly*, 24(4), 433–444.
- Becker, B. E., Huselid, M. A., & Ulrich, D. (2001). *The HR scorecard: Linking people, strategy, and performance*. Boston, MA: Harvard Business School Press.
- Boudreau, J. W., & Ramstad, P. M. (2007). *Beyond HR: The new science of human capital*. Boston, MA: Harvard Business School Press.
- Burke, W. W. (1982). *Organization development: Principles and practices*. Glenview, IL: Scott, Foresman.
- Burke, W. W. (2011). *Organization change: Theory and practice* (3rd ed.). Thousand Oaks, CA: Sage.
- Burke, W. W., Javitch, M. J., Wacławski, J., & Church, A. H. (1997). The dynamics of midstream consulting. *Consulting Psychology Journal: Practice and Research*, 49(2), 83–95.
- Burke, W. W., & Litwin, G. H. (1992). A causal model of organizational performance and change. *Journal of Management*, 18(3), 523–545.
- Bushe, G. R., & Marshak, R. J. (Eds.). (2015). *Dialogic organization development: The theory and practice of transformational change*. Oakland, CA: Berrett-Koehler.
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9(3), 621–640.
- Church, A. H. (2001). The professionalization of organization development: The next step in an evolving field. In W. A. Pasmore & R. W. Woodman (Eds.), *Research in organizational change and development* (Vol. 13, pp. 1–42). Greenwich, CT: JAI Press.
- Church, A. H., (2003), Organization development. In J. E. Edwards, J. C. Scott, & N. S. Raju (Eds.), *The Human resources program-evaluation handbook* (pp. 322–342). Thousand Oaks, CA: Sage.
- Church, A. H. (2013). Assessing the effectiveness of talent movement within a succession planning process. In T. H. DeTuncq & L. Schmidt (Eds.), *Integrated talent management scorecards: Insights from world-class organizations on demonstrating value* (pp. 255–273). Alexandria, VA: ASTD press.
- Church, A. H. (2014). What do we know about developing leadership potential? The role of OD in strategic talent management. *OD Practitioner*, 46(3), 52–61.
- Church, A. H., & Dutta, S. (2013). The promise of big data for OD: Old wine in new bottles or the next generation of data-driven methods for change? *OD Practitioner*, 45(4), 23–31.
- Church, A. H., Ginther, N., M. Levine, R., & Rotolo, C. T. (2015). Going beyond the fix: Taking performance management to the next level. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8(1), 121–129.
- Church, A. H., Golay, L. M., Rotolo, C. T., Tuller, M. D., Shull, A. C., & Desrosiers, E. I. (2012). Without effort there can be no change: Reexamining the impact of

- survey feedback and action planning on employee attitudes. In A. B. Shani, W. A. Pasmore, & R. W. Woodman (Eds.), *Research in organizational change and development 20* (pp. 223–264). Bingley, UK: Emerald Group Publishing Limited.
- Church, A. H., & Oliver, D. H. (2006). The importance of taking action, not just sharing survey feedback. In A. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies and applications* (pp. 102–130). San Francisco, CA: Jossey-Bass.
- Church, A. H., & Rotolo, C. T. (2016). Lifting the veil: What happens when you are transparent with people about their future potential? *People & Strategy*, 39(4), 36–40.
- Church, A. H., & Silzer, R. (2016). Are we on the same wavelength? Four steps for moving from talent signals to valid talent management applications. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9(3), 645–654.
- Church, A. H., & Waclawski, J. (2001). *Designing and using organizational surveys: A seven step approach*. San Francisco, CA: Jossey-Bass.
- Church, A. H. & Waclawski, J. (2010). Take the Pepsi Challenge: Talent development at PepsiCo. In R. Silzer & B. E. Dowell (Eds.), *Strategy-driven talent management: A leadership imperative* (pp. 617–640). San Francisco, CA: Jossey-Bass.
- Cummings, T. G., & Worley, C. G. (2015). *Organization development and change* (10th ed.). Stamford, CT: Cengage Learning.
- DeTuncq, T. H., & Schmidt, L. (Eds.), (2013). *Integrated talent management scorecards: Insights from world-class organizations on demonstrating value*. Alexandria, VA: ASTD press.
- Edwards, J., E., Scott, J. C., & Raju N. S. (Eds.) (2003). *The human resources program evaluation handbook*. Thousand Oaks, CA: Sage.
- Holton, E. E. III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5–21
- Huff, D., (1993). *How to lie with statistics*. New York, NY: W. W. Norton & Company, Inc.
- Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York, NY: John Wiley.
- Kirkpatrick, D. L. (1998). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler.
- Lawler, E. E. III, (2014). Sustainable effectiveness and organization development: Beyond the triple bottom line. *OD Practitioner*, 46(4), 65–67.
- Martineau, J. W., & Preskill, H. (2002). Evaluating the impact of organization development interventions. In J. Waclawski & A.H. Church, (Eds.), *Organization development: A data-driven approach to organizational change* (pp. 286–301). San Francisco, CA: Jossey-Bass.
- Nadler, D. A. (1977). *Feedback and organization development: using databased methods*. Reading, MA: Addison-Wesley.
- Phillips, J. J. (1991). *Handbook of training evaluation and measurement methods* (2nd ed.). Boston, MA: Butterworth-Heinemann.
- Roberts, D.R., & Robertson, P. J. (1993). Positive-findings bias, and measuring methodological rigor, in evaluations of organization development. *Journal of Applied Psychology*, 77(6), 918–925.
- Rothwell, W. J., Sullivan, R., & McLean, G. N. (Eds.). (1995). *Practicing organization development: A guide for consultants*. San Francisco, CA: Jossey-Bass/Pfeiffer.
- Rotolo, C. T., & Church, A. H. (2015). Big data recommendations for industrial-organizational psychology: Are we in Whoville? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8(4), 515–520.
- Savitz, A. W., & Weber, K. (2013). *Talent, transformation, and the triple bottom line: How companies can leverage human resources to achieve sustainable growth*. San Francisco, CA: Jossey-Bass.
- Shull, A. C., Church, A. H., & Burke, W. W. (2014). Something old, something new: Research findings on the practice and values of OD. *OD Practitioner*, 46(4), 23–30.
- Svyantek, D. J., O'Connell, M. S., & Baumgardner, T. S. (1992). Applications of Bayesian Methods to OD evaluation and decision making. *Human Relations*, 45(6), 621–636.
- Terpstra, D. E. (1981). Relationship between methodological rigor and reported outcomes in organization development evaluation research. *Journal of Applied Psychology*, 66(5), 541–543.
- Waclawski, J., & Church, A. H. (2002). Introduction and overview of organization development as a data-driven approach for organizational change. In J. Waclawski & A. H. Church, (Eds.), *Organization development: A data-driven approach to organizational change* (pp. 3–26). San Francisco, CA: Jossey-Bass.
- Woodman, R. W., & Wayne, S. J. (1985). An investigation of positive findings bias in evaluation of organization development interventions. *Academy of Management Journal*, 28(4), 889–913.

Allan H. Church, PhD, is Senior Vice President of Global Talent Assessment & Development at PepsiCo. Over the past 16 years he has held a variety of roles in organization development and talent management in the company. Previously he was with Warner Burke Associates for almost a decade, and before that at IBM. He is currently on the Board of Directors of HRPS, the Conference Board's Council of Talent Management, an Adjunct Professor at Columbia University, and Associate Editor of *JABS*. He is a former Chair of the Mayflower Group. Church received his PhD in Organizational Psychology from Columbia University, and is a Fellow of SIOP, APA, and APS. He can be reached at Allan.Church@pepsico.com.

Copyright of OD Practitioner is the property of Organization Development Network and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.