

QUANTIFYING QUALITATIVE OD RESULTS: DISPELLING THE TOUCHY-FEELY STIGMA

Lisa Nielsen

310 Orchid Drive

San Rafael, CA 94903

Email: redesignod@gmail.com

ABSTRACT

Evaluation enables agility and ability to self-correct mid-course. It provides Organization Development (OD) practitioners with an opportunity to continually learn from their experiences to catalyzing the evolution and refinement of their tools and skills. Oddly, the field of OD does not have an extensive history of evaluation; the quality, scarcity and validity of its evaluations have been heavily questioned since the field's inception. In contrast to the reflective rigor one might expect from a discipline that advocates high degrees of reflection from its clients, OD evaluations have provided largely anecdotal information; criteria for success has been subjective and testimonial in nature. This phenomenological research paper examines what forms of evaluation current OD practitioners utilize, what is missing from their approach, and the implications this has for the longevity of the field of OD. The ideological lineage of Behavioral Science epistemology leads to encouraging a type of empirically based evaluation practice, as represented in Campbell and Stanley's (1963) 12 Threats to Validity that may not be a good fit for Behavioral Science consultants engaging in systemically scaled interventions in a business context. This paper makes recommendations for further developments in evaluation methodology that more aptly suit a capitalist marketplace.

INTRODUCTION

The field of Organization Development (OD) was founded upon aspirations of innovation and improvement. In the spirit of continual improvement, one would anticipate the field of OD to have an established and reliable method to evaluate the effectiveness of its technologies and interventions. As Lewin (1946) wrote:

If we have no criteria for evaluating the relation between effort and achievement, there is nothing to prevent us from making the wrong conclusion and to encourage the wrong work habits. Realistic fact-finding and evaluation is a prerequisite for any learning. (p. 35)

Evaluation provides an opportunity for OD practitioners to continually learn from their experiences and to catalyze the evolution and refinement of their tools and skills. Oddly, the field of OD does not have an extensive history of evaluation; the quality, scarcity and validity of its evaluations have been criticized since the field's inception by authors such as Bennis (1965), Dunnette (1974), Lippitt, (1969), Morrison (1978), and Pearse (1970). In contrast to the reflective rigor one might expect from a discipline that asks such reflection from its clients, OD

Quantifying OD Results

evaluations have provided largely anecdotal information; criteria for success has been subjective and testimonial in nature (Hartke & Vicars, 1984).

Relevancy of this Research

The ability to understand cause and effect (single loop-learning) and to reflect upon and evaluate the underlying assumptions guiding those causal relationships (double-loop learning), is necessary if practitioners and organizations are to make informed decisions in rapidly changing and uncertain contexts (Argyris & Schon, 1974; 1978; Argyris, 1982; 1990).

There is a common perception in the business world of OD as being a soft skill, i.e. a communication, conflict resolution, or teamwork tool (Business Dictionary, 2010). While nice, soft skills are not regarded as critical to the bottom line; especially in the recent economic environment which has required of organizations unprecedented levels of innovation and financial prudence (Marshak, 2005). Many OD practitioners would argue that soft skills *are* critical to the bottom line, but at present, “executives ignore OD or relegate it to the bowels of the organization (Bradford & Burke, 2005, p. 7). To reach the level where budgets are drawn and strategies planned, OD practitioners must learn to translate results into “the language and how profit is made and costs contained according to various business models” (p. 8). Otherwise the tremendous value of OD work remains unseen, unheard, and undervalued.

Purpose of This Research

The purpose of this research is to discover what types of OD intervention evaluations are currently being utilized in the field of OD and what avenues for further development may be useful. The research also briefly explores why evaluation has not become standard procedure in OD and what the arguments against evaluation might be. This research addresses the OD field’s ability to sustain itself in a changing business marketplace.

LITERATURE REVIEW

Despite Lewin’s urging for evaluation in 1946, published literature indicates that early attempts to conduct OD intervention evaluations were not highly successful and received considerable criticism. According to Vicars and Hartke (1984):

When systematic evaluations have been attempted, for the most part they have had serious methodological weakness: Limited use of control groups, focus on short-term changes, and lack of independent evaluations are three in particular that have been complained about frequently by OD critics. (p. 177)

Vicars and Hartke’s (1984) research was inspired by an assertion by Morrison (1978) that “The studies of OD evaluations published to date do not conform to established criteria for internal and external validity for effective social science research and evaluation” (p. 65). Vicars and Hartke re-tested published OD evaluations, using Morrison’s criteria, which were based on Campbell and Stanley’s (1963) Twelve Threats to Validity, as criteria of the quality of evaluation research designs. The 12 threats are:

Quantifying OD Results

1. *History*-Effects of simultaneous events. Control group combats this threat.
2. *Maturation*- Can the results be attributed to natural development of the group?
3. *Testing*- Did pre-test and post-test sensitize respondents?
4. *Instrumentation*-Changes in the measuring instrument?
5. *Statistical Regression*- Were the study group selected based on extreme scores?
6. *Selection Bias*- Difference in experimental group and control group?
7. *Experimental Mortality*- Was there a difference in the participants between the pre and post tests?
8. *Selection-Maturation Interaction*- Is there any characteristic in the experiment group that is not present in the comparison group that would naturally lead to the changes observed?
9. *Interaction Effect of Testing*- Did the pre-test make the experiment group more sensitive to the experimental variable?
10. *Interaction Effects of Selection Biases and the Experimental Variable*- Were there characteristics of the experimental group population that would cause the results of the experiment to differ from those to be expected in the general population?
11. *Reactive Effects of Experimental Arrangements*- Were there qualities about the setting that made the results not generalizable to a non-experimental setting?
12. *Multiple Treatment Interface*- Were there other treatments applied to these groups that may have contributed to the effects noted?

The findings of Vicars and Hartke (1984) indicated that while there was a slight improvement in evaluation research design between 1978 and 1984, “Too many of these evaluations have provided only anecdotal information, and criteria for success have been too subjective and testimonial in nature” (p. 177). In an attempt to discover what was blocking successful OD evaluations Armenakis, Field, and Holley (1976) surveyed 269 members of the Organization Development Network to ascertain the evaluation challenges of OD practitioners. They discovered, “The problem most frequently encountered by these change agents was the difficulty in selecting and quantitatively measuring ‘soft’ criteria” (p. 1151). The second most frequently cited problem was the difficulty in employing comparison groups to evaluate change.

The Influence of Positivism on Evaluation Methodology

It may be useful at this juncture to examine the source of evaluation methodology applied to the OD field. Kurt Lewin is attributed with being one of the five most influential pioneers of OD (Weisbord, 2004). “The intellectual foundation for OD began in the 1940’s with the research and writing of Kurt Lewin and his protégées” (Greiner & Cummings, 2005, p. 89). Lewin is also credited as being the founder of Modern Social Psychology (NTL Institute, 2008). In 1916 Lewin obtained his Ph.D. at the University of Berlin, completing his dissertation under Dr. Carl Stumpf who was interested in empirical research methods in experimental psychology, and who became one of the pioneers in the empirical research discipline (Stumpf, 1930). In 1944, Lewin was appointed Director of MIT's newly created Research Center for Group Dynamics (Marrow, 1969). “Lewin’s life, like Taylor’s, was marked by a passion for experimentation and a faith in science” (Weisbord, 2004, p. 76). This brief history is relevant because OD evaluations, as analyzed and critiqued by the previously mentioned authors, were based on the model of empirical testing, as developed by the psychologists (a discipline that grew out of the medical field), such as Carl Stumpf in the 20th century.

Quantifying OD Results

Empirical testing roots trace back over a thousand years to the scientific method as originated in the time of Aristotle, further developed by Ibn al-Haytham and solidified by Galileo Galilei (Moulton & Schifferes, 1960; Kline, 1985). Much of the orientation of the scientific method can be summarized by the philosophy of positivism, a belief that only that which is based on sense experience and positive verification can be considered authentic knowledge. “The whole premise of data-based change (for example, action research and survey research methods) presumes the existence and validity of an objective, discernable reality” (Marshak, 2005, p. 25).

Since much of modern OD practice is based on post-modern thought and theories of social constructionism which stand in stark contrast to positivism, this research paper asks: Is there a methodology to evaluate OD interventions that would suit the discipline more aptly than traditional empirical testing?

METHODOLOGY

Research Design

A qualitative, phenomenological research method was used to investigate what type of evaluations current OD practitioners utilize to measure the impact of their interventions within organizations. Research participants were pursued through a purposeful selection method to form an eight person panel. All of the participants were asked the same 15 interview questions; however, differing follow-up questions were asked for clarification purposes. All interviews lasted approximately one hour.

Participant background summary

- Years experience: 5-30; average 15
- 50% have experience as internal and external consultants, 50% solely internal
- 100% for-profit employers including: Kaiser Foundation Health Plan, Motorola, Agilent (Previously HP), Fosters Estates, Implementation Management Associates, Boeing, and a “Big Four” accounting/advisory firm

Data Analysis

Analysis of the data was pursued through a coding and thematic analysis approach. The design strategy for developing substantive thematic category topics during the data collection process was based on the Action Research Model. Quotes were selected from each interview that captured the speaker’s salient points and were able to stand alone and hold meaning. These quotes were anonymously placed on a wall, and grouped into related categories to identify emerging cross-subject themes for analysis and discussion.

Research Augmentation

In addition to the phenomenological interview data collection, a ten-question anonymous online survey was designed and distributed to alumni and faculty of the Sonoma State Organization

Quantifying OD Results

Development Masters Degree program as well as the Fabulous Facilitators professional network, which is described as a “learning community of seasoned facilitators and those new to the field,” founded in 2002. All recipients were also requested to forward the survey on to any additional professional networks. Fifty-four voluntary respondents replied.

A limitation of the anonymous survey is that results were mostly sourced from a Sonoma State University alumni pool and California Bay Area practitioners. While respondents were encouraged to forward the survey on to professional colleagues to increase diversity of the respondent pool, this common training and/or geographical proximity could influence practitioner habits and thus survey results.

RESULTS

Each participant was asked the same set of interview questions. These focused on four areas:

- 1) The respondent’s background and the level at which they typically enter organizations.
- 2) The discussion of ‘success expectations’ with new clients during the contracting phase.
- 3) Post-engagement evaluation habits.
- 4) Reflective questions regarding establishing credibility with new clients.

Initially, interviews were pursued to discover what types of evaluations were being performed by successful OD consultants, specifically chosen for their involvement in large, metric-centric corporations, with a hypothesis that these practitioners may be the pioneers who could point toward innovative evaluation technologies. However, it was discovered during the interviews that evaluation practices, even amongst this esteemed coterie, are scarce or subjective in nature. The major themes that arose from the interviews are:

- ❖ What evaluation methods are being utilized?
 - Existing metrics
 - Tracking milestones
- ❖ What gets in the way?
 - Evaluation is complicated, time consuming and expensive.
 - It’s the client’s job.
- ❖ What is done instead?
 - Demonstrate business acumen and language
 - Form “partnerships” with clients

Existing Metrics

Often times, the first thing respondents identified when speaking of methods used to evaluate the results of their interventions was reliance upon pre-existing metrics within the organization such as sales figures, employee satisfaction survey results, inventory numbers, employee retention, attrition rates, leadership scores, and cycle-times. While not all organizations track figures, most respondents felt that “ideally the client has a set of metrics in place we can use” (R#3). *From here forth, ‘Respondent’ will be shortened to “R.”* Utilizing existing metrics was described to be ideal because it eliminates designing and conducting an extensive evaluation process, and secondly because if the client is already tracking these figures, the consultants presume these figures are important to the client and more likely to be meaningful.

Quantifying OD Results

Tracking Milestones

Interview respondents stated that they often rely upon achievement of key milestones or deliverables to track the success of a project. Sometimes “evaluation” is simply a question of, ‘Did it get accomplished?’ Commonly indicated milestones include “typical life-cycle stages: design phase, delivery, implementation, modification”(R#7) and examining whether these stages were reached and completed.

Evaluation is Complicated, Time Consuming and Expensive

Frequently, when asked about evaluation techniques, interviewees instead responded with reasons evaluation does not always happen. All of the interviewees shared the observation of R#2 that “One of the problems with measurement is that there are so many other contributing factors.” The other often cited challenges to evaluation were time and money restraints, “People are just not going to spend a lot of money on measurement because senior leaders are all about installation and then on to the next thing. They get reinforcement for installation. Strategy design is much sexier than measurement”(R#6).

It’s the Client’s Job

An additional theory was offered by several interviewees that evaluation of the effectiveness of OD interventions is “nearly impossible because I’ve stepped into what they [the client] want to take credit for” (R#4). Not all interviewees indicated that they experienced evaluation resistance from the client. Many respondents, however—particularly the external consultants—expressed that the evaluation is the responsibility of the client; “they handle that part internally” (R#7).

Demonstrate Business Acumen and Language

Since evaluation was largely described to be inconsistent or problematic, what happens instead? How do consultants establish credibility with their clients? Many of the interviewees described their ability to establish rapport through their initial scoping conversations, demonstrating to potential clients that they understood their business concern or impediment and were able to knowledgably converse about solutions. The general sentiment of the interviewees was, “I’ve got to be able to speak their [clients’] language and understand what their issues are. I don’t need to be a technical expert, but I need familiarity” (R#8).

Form Partnerships with Clients








Another factor often described by respondents was the propensity to continually check-in with the client throughout the engagement asking “is this on target?” and tracking success by relying upon the client’s subjective sense of the project’s usefulness. “As we build the process, it is checking back in with the client, ‘Is this working?’ ‘Is this going to get you where you want to be?’ ‘Is this what you want to be doing?’ A lot of it is constant renegotiation and communication because the objective changes” (R#5). Almost every interviewee echoed this sentiment: “If I start with, ‘how can I help you be successful?’ people want to work with that” (R#1).

Quantifying OD Results

Research Augmentation Results

Much of what was discovered during the interviews also surfaced in the anonymous survey results. Seventeen percent of the respondents reported that the majority of their work has been as an internal consultant. Fifty-nine percent report that the majority of their consulting work has been external and 24% stated that they have consulted in a combination of internal and external roles. Table 1 demonstrates the tenure of survey respondents.

Table 1. Survey responses to “How long have you been working as a practitioner?”

		Response Percent	Response Count
0-3 years		24.1%	13
4-5 years		11.1%	6
6-10 years		24.1%	13
11-15 years		16.7%	9
16-20 years		5.6%	3
21-25 years		11.1%	6
26-30 years		0.0%	0
31+ years		7.4%	4
answered question			54

When anonymous survey respondents were asked how frequently they discuss expectations for success with their clients during the contracting phase of their engagements, 63% reported that they always discuss expectations for success; however only 13% responded that these discussions include hard data (see Figure 1).

Quantifying OD Results

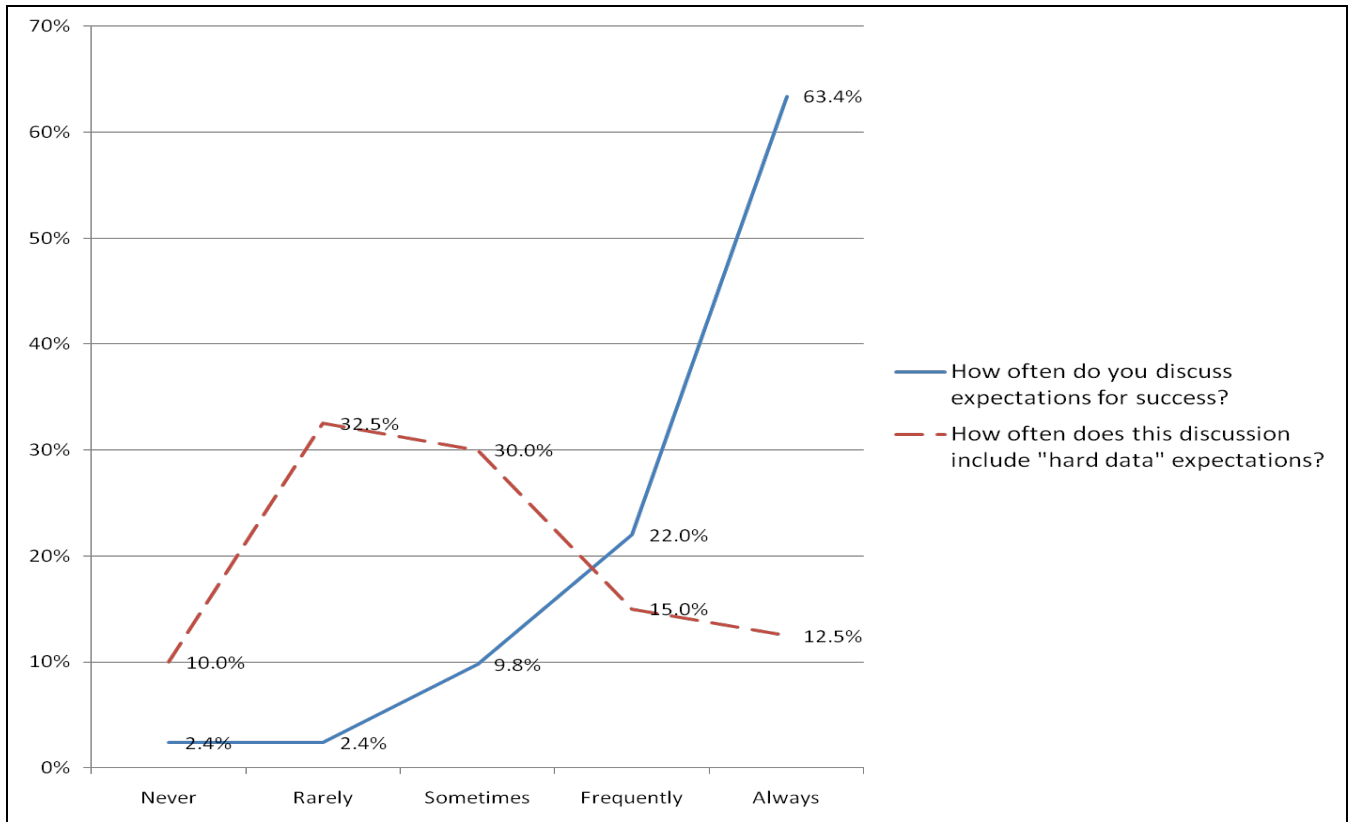


Figure 1. Discussing Expectations for Success During Contracting

When asked what method practitioners use to evaluate the results of their projects, the most frequently used form of analysis was anecdotal conversations (see Figure 2).

Quantifying OD Results

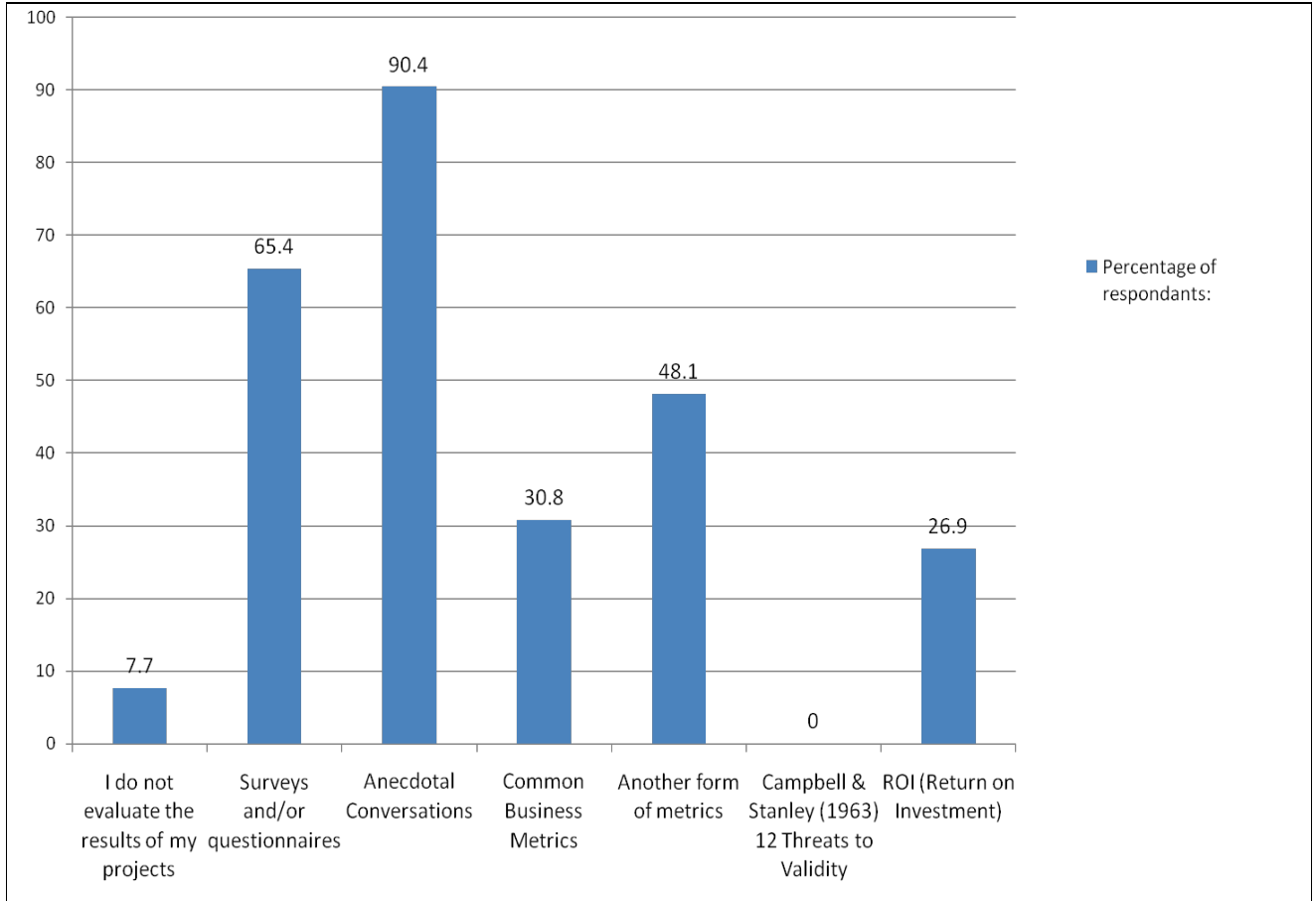


Figure 2. Survey responses to “What methods (check all that apply) do you use to evaluate the results of your engagements?”

When asked how practitioners establish credibility with new clients, only 11% believed having financial figures or metrics of previous engagements is essential to establishing credibility with new clients (see Figure 3).

Quantifying OD Results

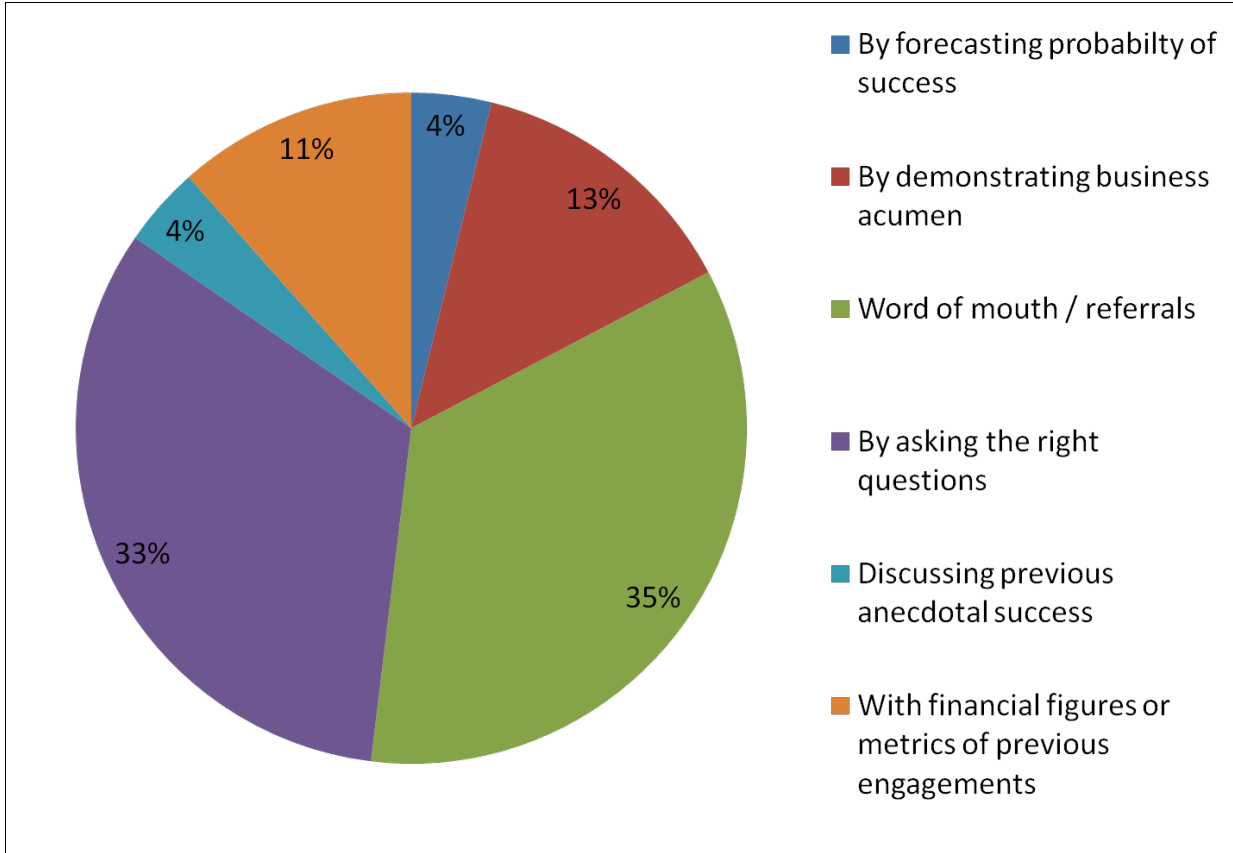


Figure 3. Survey responses to, “How do you establish credibility when contracting with a new client?”

Thirty-one respondents replied to the open-ended question, “How do you know the difference between a successful intervention/engagement and an unsuccessful one?” Answers were coded into three categories of: clearly subjective, unclear, and clearly objective (see Table 2).

Table 2. Coding of Anonymous Survey Results

Quantifying OD Results

Q9. How do you know the different between a successful intervention/engagement and an unsuccessful one?	Clearly subjective	Unclear	Clearly objective
2/24/11 12:22AM : The client and I define what will be considered successful up front (.g., improved communication among team members). If the criteria is met then it was a successful project.			
2/24/11 12:26AM: Whether the customer is happy or not.			
2/24/11 12:32AM: clients' responses & feedback			
2/24/11 12:34AM: By the commitment of the leader around what it is that he/she wants to cause.			
2/24/11 12:36AM: success will realize the projected outcomes			
2/24/11 12:36AM: 1) consistent and sustainable follow through of plans upon completion of business meeting 2) Candid and transparent dialogue with business manager/operator 3) Full participation and engagement			
2/24/11 1:26AM: Situational- Did you meet the agreed upon objectives for the engagement?			
2/24/11 1:48AM: good evaluations, call back for more work, attainment of stated goals, constant communication			
2/24/11 3:09AM: Client's definition combined with internal department metrics			
2/24/11 3:12AM: The clients expectations and outcomes have been met. Occasionally there will be a report with a clear ROI business case.			
2/24/11 3:35AM: It can be difficult to tell. Mostly, anecdotal from the client. Also, if they re-engage me.			
2/24/11 4:27AM: Client satisfaction. Follow-up later. Referrals			
2/24/11 4:28AM: In discussions with the client manager. Also, for #8, I would say that Listening to what the client has to say is actually #1 for me in establishing my credibility.			
2/24/11 4:57AM: Client is empowered and equipped to build on their success			
2/24/11 6:06AM: the quality of change - it is palpable			
2/24/11 7:35AM: at the beginning we have outlined the outcomes for a successful project - whether or not we've met these outcomes.			
2/24/11 7:42AM: The client opens their organization to me if it's successful. The client says thank you and good bye, you've been a great help if unsuccessful.			
2/24/11 4:34PM: Difficult to tell in the immediate. I continue to keep a relationship with the organization over time to determine the effectiveness of the intervention.			
2/24/11 7:59PM: Whether the metrics identified in contracting before and during the engagement are met or exceeded determines the degree of success.			
2/25/11 4:15AM: Client perception, team competency development, impacted agent feedback on process			
2/25/11 4:29AM: Happiness of the client, but not always. Sometimes I see a change in culture or behavior because of an engagement (even if it ends in the beginning stages and does not finish) and I often see those little changes as a success. Of course, the client probably wouldn't agree, but sometimes, I'm good with that.			
2/25/11 5:36AM: If we have met the stated objectives of the engagement.			
2/25/11 5:44PM: Depends on the time horizon. Often I find that the grandest long term successes feel the most chaotic. When our engagement or contribution is "over" people generally tell you how they feel it went. If it was a big and complex change which spans a decade, and we did a 3 month intervention, I expect that some stakeholders will feel much relief, some will feel much stress, and the majority will be confused or mostly unaware of the intervention.			
2/25/11 6:26PM: It is successful when the client feels satisfied when I leave, even if we didn't achieve the original outcome. It is unsuccessful anytime the client is dissatisfied with my work. Even if I achieved contracted objectives, if the client does not feel it was successful there was a disconnect somewhere in the process.			
2/25/11 7:04PM: When people can see that they've moved toward where they want to go. The metrics/indicators for that vary from place to place. Success is in the eye of the beholder.			
2/25/11 9:25PM: By the completion of deliverables, learning objectives, capacity building. A key effort of mine is to ensure that I build the capacity of any person I work with.			
2/26/11 7:41PM: It depends on the agreed upon goal established at the beginning of the project - and it is a recap conversation toward the end of the project.			
2/28/11 2:25AM: Discussions with the client afterwards, recognizing that an organization has adopted new ways of doing things, the ideas persist and they tell me they are pleased.			
2/28/11 4:05AM: by the level of energy when the last intervention is done, by debrief with client manager and by status quo on the results in forms of action planning and other tangible outcomes.			
2/28/11 11:03PM: Was the project implemented? Does the client think the project was a success? Were their metrics? Will/can I use the client as a reference?			
3/6/11 4:52PM: Business results. If the end state is not better than the beginning state, then no change has occurred. If that's the case, the intervention was unsuccessful.			

Quantifying OD Results

DISCUSSION

As evidenced in the data, the majority of practitioner respondents rely upon anecdotal conversations and subjective surveys as evaluative measures, strikingly similar to the research findings of Vicars and Hartke (1984). While this data pool is limited and not considered to be representative of the entire population of OD practitioners, the data is consistent with current criticisms and concerns of OD practices. The apparent lack of objective analysis or hard data reveals how OD acquired the stigma of being “touchy-feely” and/or “soft.” As Robert Marshak (2005) writes, “Given the language, values, and ideological orientation of OD, it should not be surprising that it has been labeled by its critics as ‘too touchy-feely’ ” (p. 25).

In addition to harming credibility, not evaluating results also inhibits an opportunity to engage in double-loop learning about the effectiveness of OD interventions. While OD practitioners often report forming thought-partnerships with their clients to inquire, “Is this it; are we getting it?” which is a form of reflection, it exists within a dangerously limited realm of subjectivity and personal experience.

Subjective Analysis vs. Objective Analysis

The severe limitations of subjective analysis are troubling when one looks at the responses to the anonymous survey question, “How do you know the difference between a successful intervention/engagement and an unsuccessful one?” Out of 31 responses (See Table 2) ten responses (32%) were unclear and only two (6%) were objective. The bulk of responses (61%) relied upon the subjective perceptions of the client, citing answers such as:

- “Client satisfaction.”
- “It can be difficult to tell. Mostly anecdotal from the client.”
- “The quality of change- it is palpable.”
- “Happiness of the client, but not always. Sometimes I see a change in culture or behavior because of an engagement (even if it ends in the beginning stages and does not finish) and I often see those little changes as a success. Of course, the client probably wouldn't agree, but sometimes, I'm good with that.”
- “Success is in the eye of the beholder.”
- “By the level of energy when the last intervention is done.”

These very distinctly subjective analysis methods rely upon the subjective perceptions of the client, how the client or stakeholder “*feels*,” the “*level of energy*,” or “*palpable*” level of change which are not units of measurement that can be verified by third parties. Even more concerning is that the subjective responses of the client are then interpreted through the consultant’s own subjective framework. Anecdotal conversation, as is so common in OD practitioner methodology of being “in partnership” with their client, is passed through two layers of subjective interpretation.

The subjective anecdotal nature of evaluative methods by current OD practitioners may help explain why many of the interview respondents reported that reaching project milestones is often substituted for tracking achievement of pre-determined hard data results. However, this approach

Quantifying OD Results

of, ‘Did we do it?’ versus ‘How well did we do it?’ does not examine the quality or effectiveness of the intervention. “Isn’t one of the key purposes of OD consultation to help human beings see where their position may lead to ineffective action?” (Argyris, 2005, p. 118).

Inductive Reasoning on the Lack of Objective Evaluation Practices

Argyris (1990) argues that people utilize subjective defensive reasoning to protect human beings and their systems from threat. Perhaps he is correct that there is a sense of security in not being held accountable by the objective measurement of hard data. The role of client/consultant “partnership” was heavily emphasized in the interviews as a critical factor to success. A side-effect of partnership is a loophole for accountability, and perhaps consultants are not eager to eliminate that loophole.

There is also an inverse hypothesis, a conveyed sense of humbleness, which surfaced in a third of the interviews. Consultants seemed to believe the credit for that success was not theirs to take. As one interviewee remarked, “My job is to make other people look good” (R#4). Many consultants spoke of approaching their clients with an “I’m here to help *you* attitude.” This attitude illuminates the source of the theme “Evaluation is the Clients’ Job,” assumedly due to a belief that achievements should be credited to the client. Again, this may also contain a shadow side indicated by Argyris (1990) of the safety provided to the consultant in lack of ownership.

There is a third hypothesis to explain why objective evaluations are not common. Perhaps the ideological lineage of our epistemology leads to encouraging a type of empirically based evaluation practice, as represented in Campbell and Stanley’s (1963) 12 Threats to Validity, that simply is not a good fit for the practice of OD.

The problem with evaluating group interventions empirically is that organizations and teams are ubiquitously unique. The conflicts, history, member status dynamics, friendships—all the incredibly intricate ingredients that comprise group dynamics are impossible to replicate in a control group. It is impossible to attain two identical groups. It stands to reason then, that early attempts at evaluation were highly prone to criticism when held against validity threats. And because, as Behaviorists would argue, humans are not prone to maintain doing what they know isn’t working; it is no surprise that the practice of evaluation in OD engagements atrophied.

What *Can* Be Done?

A recent publishing by Morris, Storberg-Walker and McMillian (2009) advocates developing an “OD-Intervention Metric System” with the use of an applied theory building method to examine work/life balance issues. The study claims to adopt a systems view of human capital to evaluate the financial return on investments (ROI) of work/life interventions.

It is interesting to compare the standard for “metrics” versus the standard for empirical evaluation. According to Becker, Huselid and Ulrich (2001) metrics possess credibility, can be believed by a reasonable person, have meaningfulness, are legitimate and accurate, reliable and valid, and possess strategic value for the end user. Typical metric categories consist of: volume, quantity, cost, income, time, quality, stakeholder reaction, rate, ratio, and categories or levels (Morris et al., 2009).

Quantifying OD Results

Missing from this list of metric requirements is the need for control groups and absolute elimination of internal and external validity threats, as is required in empirical testing. In contrast, “The foundation of a good metrics system is measurement of attributes and relationships that measure hypothesized causality as the goal for making decisions and assessing performance” (Morris et al., 2009, p. 425). Metrics allow much more freedom and flexibility and yet their credibility is not questioned. Quite the opposite: “In business, value must be measured through metrics” (p.445).

One example of this is the common Return on Investment (ROI) measurement. While the ROI is a simplistic instrument, it is widely accepted as a useful metric to measure business operations. Validity threats are not viewed as prohibitive from conducting evaluation. Certainly it does not have the precision of empirical science, yet it provides a useful evaluation framework. While an ROI evaluation is certainly not a blanket answer, it is an indication that there may be other forms of evaluative methods than the OD community has not embraced.

Quantifying the Qualitative

As noted by the OD practitioners surveyed on the challenges of OD evaluations, the most frequently cited difficulty was quantitatively measuring “soft” criteria (Armenakis et al., 1976). Fortunately, new developments are being made in this area. In a study titled “The Role of Positivity and Connectivity in the Performance of Business Teams” researchers Losada and Heaphy (2004) coded the verbal communication of business teams, measuring the frequency of approving Positive (P) statements versus disapproving Negative (N) statements. This coding established a P/N ratio which was found to have a statistically significant correlation with the team’s performance level. The work of Losada and Heaphy (2004) is an excellent example of the quantitative research void the OD field needs to fill to establish its credibility and relevancy to the business community. As Beer (1976) lamented, “OD suffers from an inability to demonstrate the relevance of its interventions. We desperately need research aimed at developing a technology for quantifying, in organizationally relevant terms (profits, services, good will, market competitiveness, innovations, etc.), the results of changes created” (p. 50).

Systemic Implications

Organizations exist within a larger environment that impacts their well-being and ability to survive. Shifting environmental conditions are identified through an active feedback loop filtering between the organization and its environment that measures sales, buying trends, customer satisfaction, available resources, competition, etc. Organizations that choose to ignore these shifting environments tend to lack the ability to adapt and survive (Pasmore, 1988). The field of OD is not above or apart from these external environmental realities.

Implications for the Field of OD

It is clear that the field of OD can strengthen its presence in several ways. First, during client engagements, determining a mutually agreed upon goal and means of distinguishing and measuring traction, is essential before beginning the project. Ideally, this would be a metric that can be measured objectively, drawing from the metric categories listed above or inspired by a SMART goal: specific, measurable, attainable, relevant and time-bound.

Quantifying OD Results

This shift in contracting behavior requires spending time to define qualitative aspirations. Often success goals are established such as “increased communication,” or “reduced conflict.” However, objective parameters need to be specified to measure *when* communication has increased and *how much* conflict has been reduced. “In a very basic way, every project is about reconciling the fundamental polarity between the world of What-Is-Needed and the world of What-Can-Be-Built” (Conklin, 2006, p. 16). What-is-Needed might be energized innovation or an improved level of employee engagement- but these are aspirations that have no ceiling for achievement. A conversation to balance these what-is-needed items must be accompanied with an honest what-can-be-built perspective. “Energized innovation” might be objectively tracked by the number of employee contributions, engineering design submissions, item-to-market ratios, or number of new products released to market. “Employee engagement” might track attrition levels, employee satisfaction surveys, or numbers of colleague referrals. Achieving these identified what-can-be-built targets will not mean that the larger what-is-needed issues will have been “solved” but it does provide a means to objectively reflect on methodologies and success indicators undertaken by the consultant. Ideally these identified metrics will provide the client a chance to evaluate how their financial investment in the OD project is impacting their organization’s profitability.

CONCLUSION

The philosophical humanistic underpinnings of OD such as emphasis on self-awareness, communication, authenticity, inclusion and collaboration bring a perspective into the workplace that enriches lives and serves to counter-balance the profit-centered approach of the modern economy. However, if we cannot adapt to translate these contributions into a language the current external environments understands and values, the reach and impact of OD will decline. As Bradford and Burke (2005) write, “The consequences are that they [OD practitioners] will not be at the table where significant change decisions are made. Instead, at that table will be representatives from the major consulting firms. Only the crumbs will be left for the OD practitioner” (p. 213). OD practitioners need to pay attention to this risk and bring it into the collective narrative.

The OD field’s lack of adaptation to the marketplace’s orientation indicates there is a critical feedback loop in our own systemic model that is dangerously underdeveloped. We are not listening and adapting to the larger community upon whose resources we rely for survival. We need to figure out how to capture the qualitative shifts that occur during an OD intervention and translate them into quantitative examples. It is self-sabotage to remain elitist in our language and our values. The client and the greater community’s values, such as profit, must enter into our thinking and design if this field is to stay relevant and not become extinct.

The feedback loop needs to be repaired and strengthened. This discussion needs to enter our conferences and professional gatherings. Contemporary evaluation instruments need to enter our research focus. Business acumen and results evaluation need to be incorporated into graduate programs’ curriculum, and finally, as practitioners, we need to strive for dedicated personal accountability during our own engagements.

Quantifying OD Results

REFERENCES

- Argyris, C. (1990) *Overcoming Organizational Defenses. Facilitating organizational learning*. Boston: Allyn and Bacon.
- Argyris, C. and Schön, D. (1974) *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.
- Argyris, C., & Schön, D. (1978) *Organizational learning: A theory of action perspective*. Reading, Mass: Addison Wesley.
- Armenakis, A., Feild, H., & Holley, W. (1976). Guidelines for overcoming empirically identified evaluation problems of organizational development change agents. *Human Relations*, 29(12), 1147-1161. doi:10.1177/001872677602901204
- Becker, B. E., Huselid, M.A., & Ulrich, D. (2001). *HR Scored: Linking People, strategy, and performance*. Boston: Harvard Business School Press.
- Bennis, W. (1965). Theory and method in applying behavioral science to planned organization change. *Journal of Applied Behavioral Science*, 1, 411-428.
- Beer, M. (1976). On gaining influence and power for OD. *Journal of Applied Behavioral Science*, 12 (1), 44-51.
- Bradford, D. & Burke, W. (2005). *Reinventing Organization Development*. San Francisco: John Wiley & Sons.
- Conklin, J. (2006). *Dialogue Mapping*. John Wiley and Sons.
- Dunnette, M. D. (1974). Mishmash, mush, and milestones in organizational psychology: 1974. In H. Metzger (chair), *Humanizing Organizational Psychology in 1974*. Symposium presented at the meeting of the American Psychological Association, New Orleans.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2, 34-46.
- Lippitt, G. (1969). *Organizational renewal*. New York: Appleton-Century-Crofts.
- Morris, M.L., Storberg-Walker, J. & McMillian, H.S. (2009). Developing an OD-Intervention metric system with the use of applied theory-building methodology: A work/life intervention example. *Human Resource Development Quarterly*, 20: 419-449. doi: 10.1002/hrdq.20026
- Morrison, P. (1978). Evaluation in OD: A Review and an Assessment. *Group & Organization Studies (pre-1986)*, 3(1), 42-70.
- Moulton, F. R. & Schifferes, J. J. (eds., Second Edition) (1960), *The Autobiography of Science*. Doubleday.
- NTL Institute. (2008). *NTL: History*. Retrieved from <http://www.ntl.org/inner.asp?id=178&category=2>
- Kline, M. (1985). *Mathematics for the Nonmathematician*. Courier Dover Publications.
- Pasmore, W. (1988). *Designing effective organizations: the sociotechnical systems perspective*. New York: Wiley.
- Pearse, R. F. (1970). Organization development. In R.F. Moore (Ed.), *AMA management handbook*. New York: American Management Association.
- Stumpf, C. (1930). Autobiography of Carl Stumpf. In C. Murchison (Ed.), *History of psychology in autobiography* (Vol. 1, pp. 389-441). Worcester, MA: Clark University Press.
- Vicars, W., & Hartke, D. (1984). Evaluating OD evaluations: A status report. *Group & Organization Studies*, 9(2), 177-188. doi:10.1177/105960118400900203.
- Weisbord, M. (2004). *Productive workplaces revisited: Dignity, meaning and community in the 21st century*. San Francisco: Jossey-Bass/Wiley