

MEASUREMENT ISSUES IN USING PAPER AND PENCIL INSTRUMENTS: CONTRASTING THE VIEWS OF ACADEMIC RESEARCHERS, TEACHERS, AND OD CONSULTANTS

Michael R. Manning

New Mexico State University

Conrad N. Jackson

University of Alabama in Huntsville

Valaya Pathi

Tiffin University

This paper examines the validity standards applied to paper and pencil questionnaires by four social science roles: academic researchers, expert consultants, teachers, and organization development (OD) process consultants. These roles, influenced by their key stakeholders (e.g., clients, reviewers, students), employ different validity standards and socialize firm beliefs about these standards. The resulting divergence of attitudes toward measurement inhibits learning from other roles and slows social science advancement. This paper examines the assumption that each role must apply the same validity standards.

The social scientist is trained to think that he does not know all the answers. The social scientist is not trained to realize that he does not know all the questions. And that is why his social influence is not unfailingly constructive (Cronbach, 1975, p. 13).

The development of scientific knowledge is intertwined with progress in measuring constructs. The tools and instruments needed for measurement of physical phenomena are developed with relative ease by comparison to measurement of organizations and social systems. In part, this may be because of the extreme complexity and abstractness of social sciences. Further, physical phenomena can be manipulated in the laboratory with fewer of the ethical constraints which confine social and behavioral sciences. Also, physical sciences attract enormous levels of government funding compared to behavioral sciences, leading to larger and more extensive investigations.

Notwithstanding this lot, social scientists have tried to standardize their measurement processes since the dawn of the Renaissance. Huarte in the sixteenth and La Chambre and Thomasius in the seventeenth centuries attempted to develop "disciplines of assessment" in "the study of people and character" (Reynolds &

Wilson, 1985). This tradition continued notably in the works of Malthus in economics in the eighteenth century, and more recently with Marx in sociology and Cattell in psychology (Cattell & Johnson, 1986).

Standardization is important for knowledge to be cumulative and accretive—for the study of social phenomena to become a science. Much progress has been made in the academic community, it seems, toward forming consensus on measurement standards of validity and reliability in the social sciences. For example, industrial psychology, education, sociology, organizational behavior, management, and business policy graduate students often take the same methods courses using the same textbooks (e.g., Kerlinger, 1986). But not all areas of social study and practice have adopted and used these rules equally.

Validity Challenges in the Organization Development Discipline

In particular, many of the instruments widely used in management and organization development (OD) frequently come under severe criticism (King, Sherwood, & Manning, 1978). Instruments used by practitioners—managers, consultants, and facilitators—are often not tested for reliability or validity in the same painstaking detail as the test instruments developed and used by scientists. For example, the Job Diagnostic Survey (Hackman & Oldham, 1980), the Myers-Briggs Type Indicator (Myers, 1962), and the Learning Style Inventory (Kolb, 1976) are widely used in OD and training, though many researchers and theorists object to the use of these instruments, citing issues of measurement (see, for example Freedman & Stumpf, 1978; Freedman & Stumpf, 1980).

Further, the concept of an "action science" (Argyris, Putnam, & Smith, 1985) is not recognized and understood in many academic disciplines. Hence OD academicians, who integrate clinical insights with more controlled studies, can be seen as "anti-scientific," especially by academicians or clients who are driven by the assumption that logical, positivist values are science. Thus the field of OD is often criticized for measurement and evaluation which does not meet the standards imposed by others.

Is the use of these instruments in OD due simply to lower concern for the measurement issues or is it *prima facie* evidence of incompetence or intellectual laggardness? Or are the standards demanded in research settings not fully applicable in more applied settings related to organizational change? This paper will explore these issues by clarifying the measurement standards in various social science roles (i.e., academic researcher, expert consultant, teacher, and process consultant) and explore how stakeholders of these roles (e.g., clients, research subjects, managers, etc.) influence measurement standards. In particular, we will focus on paper and pencil self-report instruments, since they are widely employed in a variety of settings serving a variety of purposes.

Four Categories of Questionnaire Uses

To examine the use of paper and pencil instruments, four broad categories or roles of use for these instruments are proposed: academic researcher, expert consultant, teacher, and process consultant. The following hypothetical scenarios help convey the important issues unique to each role and its use of paper and pencil instruments.

Academic Researcher. Professor Judy Abbas, a cognitive scientist, is studying risk seeking and risk aversive behavior. Her thesis is that human resource managers are more risk seeking than marketing executives, who in turn are more risk seeking than accountants. These propositions are based on her earlier research, as well as other literature in the field. She plans to use a paper and pencil instrument to measure the risk seeking and risk aversive behavior. The measure will be constructed on the foundation of existing knowledge and on Professor Abbas's expertise. It will also be discussed with and reviewed by other experts in the field of choice behavior. Then the instrument will be tested for reliability, validity, and clarity through pilot testing, item analysis, cross-validation, and reliability analysis. The sample has been chosen with great care to provide the requisite demographic representation. Completion of the study is expected in one year.

Expert Consultant. Bill Burton is a free-lance consultant who has been employed by Beta Corporation to assess the teamwork skills of its managers. Mr. Burton holds a masters degree in industrial psychology and has worked for ten years in the personnel department of a large corporation. Beta is downsizing and reorganizing, with more emphasis to be placed on participative management and team approaches. Based on Burton's assessment, Beta intends to assign some employees to key work teams and to relocate or fire others who cannot fit into this new management scheme. Mr. Burton, as part of his assessment, will use a questionnaire he has written on employees' propensity toward the risky-shift phenomenon. This questionnaire is based on his ten years of managerial experience in industry and eight additional years of clinical experience as a management consultant. Though no formal validation studies have been performed, his use of this instrument in other organizational settings has met with general acceptance.

Teacher. Professor Cheyney teaches an organizational behavior class of 40 students at a university. The course includes a module on choice-making. Two paper and pencil questionnaires are used in this teaching module, one on "risky-shift" drawn from literature and another an amalgam of research questions investigated by various cognitive researchers. The students complete the instruments before they do their reading assignment on choice-making. The class discussion begins with students' reactions to the questionnaires they have completed. The theory and current research on risk orientation are discussed in the context of the choices students made on the questionnaire items. Professor Cheyney believes that this instruction method generates an intense student interest in the topic. He also uses a similar procedure when conducting training workshops for managers. Each

student/trainee keeps her or his completed questionnaire and results. Professor Cheyney makes no effort to aggregate individual data for research purposes.

Process Consultant. Joe Duncan is a professor of OD in the business college at a major university. He has contracted with MEGA corporation to assist with the implementation of concurrent engineering via cross-functional work teams. One of MEGA's concerns is how risk-averse engineers will be able to work in teams with risk-prone marketing personnel. Professor Duncan's approach to this situation, as is typical of OD consultants, is to facilitate the client's own decision and planning processes rather than to simply provide an expert solution which the client might have difficulty implementing. He plans to provoke team members' introspection with respect to risk taking to facilitate teams achieving an explicit consensus on this issue. Professor Duncan will ask each team member to complete the Risk Assessment Inventory for Team Members (RAITM), an instrument he constructed specifically for this client based on his clinical impression of the risk-taking dynamics teams will face. The RAITM has not been validated empirically. Individuals calculate their own scores by summing the responses from Likert scale items. Individuals will be encouraged (but not required) by Duncan to share their scores and reactions to questionnaire items. Each team will be encouraged to discuss risk taking in their group. RAITM scores will not be used for any other purpose.

While each of these scenarios features a protagonist who uses a questionnaire focusing on aspects of risky choice, the situations and objectives differ markedly, and the potential measurement issues surrounding the use of paper and pencil instruments are quite different. The significant questions we address here are: (1) Should the standards demanded of the instruments used by Abbas, Burton, Cheyney, and Duncan be similar? (2) If the same standards are required for all four scenarios, how can their obviously dissimilar objectives, settings, and constraints be served? (3) Or, if a single standard is not adequate to cover these diverse settings, how does one assure quality, consistency, and utility for each type of setting?

These four professionals construct their world views quite differently. Each is affected by a different set of stakeholders, and each must be responsive to the criteria these stakeholders use to judge research or measurement. To some extent, these four world views do overlap, though there is little communication among our protagonists. Potentially there may be much for each to learn from the others. This learning is inhibited, however, when these four worlds socialize their members toward different stakeholders and standards, and thereby judge each other's work as inadequate. We will therefore examine these four perspectives to identify where they differ and where they may be able to find agreement.

Purpose of Measurement

Obvious uses of questionnaires include measuring the occurrence of phenomena and testing hypotheses which advance theories in the field of study. This is why Professor Abbas employed her risk-seeking and risk-aversion questionnaire.

The focal point for academic researchers like Abbas is the testing of conceptual propositions under their investigation. Their questionnaires must yield measures which exhibit construct validity (Campbell & Fiske, 1959), as well as discriminant and convergent validity.

However, other purposes for using questionnaires exist in the scenarios above. Burton's risky shift questionnaire provides data for personnel decisions in a specific organization. He is not concerned that his instrument may yield results inconsistent with or unrelated to any other concepts in the body of scientific knowledge. He simply wants to ensure that his questionnaire can be used to select employees who will work successfully in teams. Mr. Burton is referred to as an "expert consultant" because he will assess the situation and make an expert recommendation to the company on hiring decisions. To provide useful advice, Burton's questionnaire must demonstrate incremental validity (Cascio, 1982). In short, it must result in hiring, promoting, or firing the right people at a rate which exceeds the correct decisions which could be made without the use of this questionnaire. Further, legally such measures are considered tests, and if their use results in a disparate impact on protected minorities, the employer must be able to demonstrate that the test is a valid predictor of the performance criterion rather than simply a socially-biased tool of discrimination (Cascio, 1982).

Professor Cheyney, the teacher, uses questionnaires to illustrate how one might operationalize an abstract cognitive or behavioral concept and to pique students' interest in social science concepts they read about. Cheyney also tries to provoke students' introspection concerning risk and choice making.

Duncan, the OD process consultant, uses a self-report questionnaire to help the client (the team members) explore hypotheses about risk. Duncan will facilitate both a self exploration and a group discussion of risk taking expectations for their team. Duncan does not consider the RAITM to be a measure in the same way that Abbas views a measure. Rather, Duncan uses data from his instrument as an *indicator* providing a provisional agenda or heuristic framework that begins/continues a personal/organizational change process. He will forewarn the client not to use these data to label people or situations nor to address issues beyond the immediate situation. Rather, these data serve to "prime the pump" to help focus and challenge individuals to look deeper at a phenomenon without arousing psychological defenses that prevent further examination (see Bowen & Jackson, 1985).

Evaluating Validity

Each of these cases may require different standards for judging the adequacy of the measurement method. Various procedures have been developed to assess the validity of a measure in a given setting. We propose five characteristics of paper and pencil instruments to illustrate our proposition that differing standards of measurement are used in each scenario.

1. Reliability—the extent to which a measure is repeatable and thus stable when there is no real change in the phenomenon being measured;

2. Convergence—the extent to which this measure of a concept converges (or correlates) with other measures of that same concept;

3. Prediction—the extent to which the measure proves to be an accurate predictor of some outcome which is of interest to the stakeholder;

4. Face Validity—the extent to which a person who is not an expert in the subject area would be able to ascertain that the measure appears to relate to the concept which the researcher/teacher/consultant intends it to measure;

5. Relevance to the Employee—the extent to which the concept being measured is seen by the employee (or participant or research subject) as important and instrumental to his or her work effectiveness, job security, work enjoyment, etc., or to some such issue for which the individual has a high valence.

To initiate a discussion of the importance of these validity characteristics for the stakeholders in our five scenarios, we offer the following observations and propose some conclusions.

The Social Construction of Attitudes Toward Validity

Academic Researcher. The academic researcher judges a measure primarily by its reliability and convergence (see Table 1). In fact, these criteria are deeply imbedded in the academic research culture and are strongly driven by journal editors, reviewers, and other researchers who comprise an influential stakeholder group. Although prediction is sometimes relevant to a given study, both journal editors/reviewers and academic researchers do not usually view the lack of prediction of "on the job" behavior as a fatal flaw when judging paper and pencil instrument validity. In fact, academic researchers are often accused of lacking any concern for the importance of their work in the real world. This tolerance may arise because social science constructs rarely predict more than a small amount of variance in dependent measures (performance, individual behavior, job satisfaction, etc.), due in large part to the complexity of the human element in organizations. Thus, if an academic researcher limits his value for an instrument solely on grounds of prediction and relevance, it may be difficult for him to generate any results to justify his work. Therefore, academic researchers place emphasis on reliability and convergence, two issues over which they have more control.

Research subjects, who are also stakeholders in this socially-constructed system, place different values on validity criteria. Subjects, who may perceive little personal value from participating in a study like Abbas', care little about the measurement properties of her questionnaire. Abbas cannot inform participants a priori of the true purpose of her study because she is fearful of creating demand characteristics and experimenter expectancy effects that may bias results. Hence, subjects have no concern about (or even realization of) prediction. To the extent that subjects can glean some meaning from the questionnaire *vis-à-vis* face validity, infer-

Table 1
Attitudes of Stakeholders Toward Validity Standards of Academic Researchers, Expert Consultants, Teachers, and Process Consultants

Validity Statistics	Academic Researchers		Expert Consultant		Teacher		Process Consultant					
	Research Subjects	Editors, Reviewers and Other Researcher	Research Investigators	Research Respondents (Client's Employees)	Client (Sponsor)	Expert Consultant	Student (Trainees)	Employers of students	Teacher	Client/ Employees	Sponsor	Process Consultant
Reliability	0	++	++	0	0	++	0	0	+	0	0	+
Convergence	0	++	++	0	0	++	0	0	+	0	0	+
Prediction	0	+	+	+	++	++	+	+	+	-	+	-
Face Validity	+	-	-	++	+	0	++	++	++	++	++	++
Relevance to Respondents	+	-	-	++	+	+	++	++	++	++	++	++

++ = Extreme necessity
 + = Desirable but not necessary
 0 = Not aware of doesn't care
 - = Troublesome but not fatal

ring relevance to their jobs or lives, they may find the questionnaire somewhat interesting. However, journal editors and reviewers may have concern about measurement if scales are too transparent, allowing subjects to easily fake or bias their responses consistent with their own agendas.

Although Abbas' attitudes toward validity standards are influenced by these multiple forces, her top priority will likely be acceptance of her work by academic colleagues. This acceptance forms the basis for tenure and promotion decisions, academic job mobility, research funding, and professional leadership. Because Abbas is working with constructs that are established in the literature of psychology and organizational behavior, she needs a measure of risk aversion which will not fail to correlate with the accepted, established measures of risk aversion. She knows that to get the highest correlation between two measures, both measures must exhibit high levels of reliability. Thus, according to standards of reward and recognition by which she is judged (by stakeholders who control her career advancement), reliability and convergence are rational choices as criteria for a good measure.

Expert Consultant. Bill Burton, the consultant hired for personnel assessment, will be affected by a different set of stakeholders with quite different expectations. The stakeholders who influence Burton's social construction of a good measure are the individuals from BETA Corporation who hired him (the client/sponsor) and the research participants (client/employees) who provide the data and are subjected to the results of his work.

Burton's clients want predictions that result in sound business decisions and practices. They care little about how he arrives at his suggestions, even if this expert advice arises from intuition not demonstrably connected to a body of research literature. They do care that his consultation assists in making decisions which are important to the firm. Thus, it is the convergence of the measure with outcomes of interest to the clients which helps the consultant and his stakeholders, not the convergence of his measure with other measures in the literature. At the outset, face validity may build the clients' confidence that the measurement is relevant to the organization. The ultimate test will be whether the targeted predictions are achieved.

If the validity of Burton's work is challenged in court (another powerful stakeholder), Mr. Burton and BETA must convince the court that the selection system implemented is legally valid. Predictive validation is the safest protection against prosecution. However, as predictive validation is an expensive and time-consuming process, concurrent validation or even content validation are often deemed acceptable by courts and thus may constitute a reasonable risk by the client.

Court interpretations of Title VII of the Civil Rights Act of 1964 clearly indicate that *face validity* is not a sufficient defense for the use of a selection procedure. Industrial psychologists suggest that the safest protection against is evidence of *predictive validity*. This requires administering the questionnaire to a group of applicants or promotables, then hiring or advancing them without regard to test

scores, and finally correlating later performance with test scores to get an indication of test validity across the full range of performance and the full range of test scores. Cross validation with a second sample provides even stronger evidence of validity. Because this procedure is obviously dramatically costly and time consuming, a less intensive (and scientifically less convincing) strategy for demonstrating validity is a concurrent validity study. The instrument is administered to a group of employees who are already employed in the relevant jobs, with evidence of validity being provided if their test scores cluster on the high end of the scale.

However, in practice the courts have often accepted an even less stringent indication of test validity. A qualified expert might convince the court that the test has content validity (i.e., that in his or her expert opinion it appears likely to measure what it is purported to measure). Even this requirement is often made moot by the court's adoption of the 80% rule, which said that if a protected minority was represented at least 80% of the proportion that this minority was found in the local labor market, there was no *prima facie* evidence of discrimination, and the company could generally use whatever selection methods they chose. In this case, Burton's issue becomes what the employer believes about the validity—and more important, the *usefulness*—of his measure. BETA's judgment about Burton's instrument rests on a subjective cost/benefit assessment based on the incremental validity of the instrument or possibly on a utility analysis (Cascio, 1982). Complicating this situation is the fact that employers cannot wait forever for a consultant's recommendations. At best, long delays may make the information moot, with hiring proceeding without benefit of his input.

Client employees, who will be asked to fill out the paper and pencil measure, will wonder how they will be affected by the data they provide. If they judge the instrument to have face validity and relevance to their work, they are more likely to be cooperative and honest. These stakeholders likely do not perceive or understand issues of reliability and convergence. Burton is concerned about their perceptions because, if they become willing participants in his consultation, the odds of success go up dramatically.

Table 1 indicates that Burton knows the importance of reliability even if his client does not. It is instrumental in ensuring legal defensibility. Burton does care about face validity for its own sake, but he knows that his work may well evoke less fear among employees and thus stir up less political opposition if the client/employees perceive the relevance of his questionnaire. He is primarily driven by the client's concern for prediction, and he knows that his income and consulting reputation are on the line, dependent on his judgments related to selecting key employees.

Teaching. In classroom settings like Professor Cheyney's, paper and pencil measures are used to teach and communicate conceptual frameworks or models and to help individuals personalize these concepts. As Table 1 illustrates, the instructor is especially concerned about the relevance of his measure. This does not mean, however, that Cheyney is unconcerned about reliability, validity, prediction,

or face validity. It is simply that an instructor can better reach the student/trainee who sees the topic as instrumental.

Professor Cheyney employs his questionnaire as a mechanism to help students apply risk taking concepts to their personal and professional lives. He reasons that when students can identify, understand, and apply these concepts to their own behavior, then they can intellectually integrate this into their understanding of others. The teacher may suggest a personal exploration of how changes in risk-taking can result in differing outcomes for the individual. Subsequent experimentation by the student will be left as his/her personal decision.

Expectancy theory suggests that personal learning is strongly affected by the student's view of the relevance of the concept to his or her own life. This may explain why student stakeholders seem to hold relevance and face validity as primary standards for judging questionnaires used in teaching. If students believe that they understand (and approve of) what the questionnaire measures (face validity), they are willing to be more genuine with their answers. If the issue being measured is relevant to their lives, they will risk honesty in exchange for the insights they can gain. The prediction of subsequent behavior may also be seen favorably by students, though their natural skepticism and/or appreciation for the complexity of behavioral causation may lead them to see prediction as less important than relevance and face validity.

Reliability and the convergence of a measure with others in the academic literature are of little concern to most students. Reliability occasionally becomes an issue when a student asks: "Do I answer in terms of being at work or being at home?" The instructor generally acknowledges that people might have different issues at work than at home and that all answers are likely to be affected by recent events, today's mood, approaching deadlines, etc. Thus, the students should answer the way s/he generally behaves/feels. All of these attitudes toward measurement appear to be similar for employers and students/trainees (Table 1). Relevance and face validity are crucial because they ensure student attention and employee support.

Process Consultant. The OD consultant's role and activities with respect to paper and pencil measures can be summarized by Argyris' (1970) intervention objectives:

1. Help the client obtain valid, relevant, and necessary data;
2. Facilitate the client's free and informed choice concerning these data; and
3. Help the client develop internal commitment to implement the plan formulated from the data.

Information provided by instruments (e.g., JDI, MBTI, LSI) may be useful for this task. The OD consultant's objective is to use questionnaires to initiate a group's exploration of work methods and procedures, much as the teacher's use of instruments aims to evoke a student's exploration of a concept. This is a different goal than that of the academic researcher who is mostly interested in the *content* of what is being measured (and convergence with similar concepts). It also differs from the

expert consultant's goal of measuring to predict future behavior or performance. For the OD consultant, questionnaire data is a starting point for developing deeper contemplation and discussion to explore critical issues. It should raise the relevant constructs to the conscious level in the organizational members' minds in such a way that the constructs' relevance becomes compelling to these individuals. The participants then explore the implications of these issues to their particular work context. The skillful consultant uses this motivated inquiry to evoke constructive confrontation of critical issues that would otherwise be denied or avoided.

Both the client/sponsor and the client/employee will also see face validity as important. It will intuitively provide them with a motive to complete the questionnaire and will help ensure that they understand the insights it can provide. When the participants understand the questionnaire, they can better monitor their responses—i.e., they feel in control rather than worrying that the consultant is some sort of "shrink" who is "psyching them out." They will respond to items with greater care if it is clear that the questions relate to their own work experience. Abstract questionnaires, like abstract exercises (e.g., power simulations, ropes courses), may be resisted unless the consultant can convey their relevance and create a sense of face validity.

No questionnaire can guarantee truthful answers. The process consultant realizes that truth is not usually found in questionnaire data. But in a sufficiently "safe" environment, these data can lead to more honest sharing of information. So questionnaire results are only the beginning point on the road for the client's journey toward valid data. Thus, if questionnaire data are not reliable and do not converge with the academic literature, this is not necessarily a problem for the OD consultant. Further, although a prediction may appear to be a positive for the client/sponsor, it may prove troublesome to the client/employees—and thus to the consultant. If employees are skeptical about how results from a questionnaire might be used by others to control their destiny, they may participate in a backlash of resentment borne of fear for their own personal careers.

Confirmation of data validity in the OD consultation comes from subsequent discussions of the data which explore and elaborate on issues and individuals' reactions. So the validity of Duncan's data depends not so much on the questionnaire itself as on his skill at "processing it out" (discussing the data) with the client. As organizational members experience safety and positive results from their tentative sharing and talking about organizational issues, they feel encouraged to share and ask for deeper, more honest, and critical (and thus more threatening) feelings and thoughts about key issues. In short, simple written answers to questions do not evoke successful action policies (Jackson & Manning, in press). Consensual, face-to-face discussions, facilitated by trained process consultants, transform the original data into a more usable form, creating a new level of understanding and commitment which is more likely to result in action by the client.

To the OD consultant, therefore, validity is achieved and verified by public discussion, consensus testing, action planning, and follow up. Relevance is a key to

this process. Perceived relevance and face validity build confidence that the consultant's efforts will be linked to organizational and stakeholders' personal payoffs.

Good Enough for Consulting Work?

It is widely recognized that many instruments used for training and organizational interventions frequently do not meet the standards demanded by academic researchers. Even given their differing goals for using a questionnaire, however, is there justification for OD practitioners to ignore the measurement standards established in the organizational sciences? Arguments exist for either side of this issue.

Certainly a questionnaire that is both a measure and an indicator would have merit. However, precise and succinct measures of most group and individual behavioral/attitudinal/cognitive phenomena which are robust to most situations simply do not exist. To develop valid measures for every consulting purpose would be prohibitively costly. And given the slow pace of scientific validation of measures, much valuable applied work would have to wait for years or decades. Just as the Chairman of the Federal Reserve System cannot delay decisions on adjusting the discount rate until academic economists come to consensus, so managers must make the best decisions they can now, given what they know about the situation. Thus, even expert consultants must often use instruments despite the lack of empirical evidence of construct and predictive validity.

Moreover, problems often accompany the use of validated instruments in OD consultation. Perhaps the most unfortunate fallout would be the potential for evoking defensive reactions by test-takers. Argyris (1968) has identified several unintended consequences of rigorous research, including subject willingness to allow his or her behavior to be manipulated by the investigator or even knowingly giving incorrect answers. The psychology literature contains numerous warnings and much research evidence on research subjects' suspicions (e.g., see Cook & Perlin, 1971), demand characteristics (e.g., see Orne, 1962), and evaluation apprehension (e.g., see Adair & Schachter, 1972).

Another danger for the OD consultant using scientifically validated measures is the temptation to "make expert recommendations." Although it may seem ironic to the scientist that facts can be anything but friendly, one of the most feared situations of the OD consultant is that the client will not "own" the diagnosis (out of fear, political turf protection, or even sheer stubbornness) and thus not feel motivated to do anything about the diagnosis. Ultimately, it is not the recognition of reality in the mind of the consultant or even the boss which changes an organization but that recognition in the minds of all the people who must adapt to it. Getting to the truth, therefore, may be vital. But this will not be achieved with a questionnaire if the data are not owned by the client.

Transferability of Knowledge: Practice to Research

Does the information unearthed by predictor instruments like Burton's or indicator instruments like Cheyne's have any value in advancing our scientific knowledge of human systems? The answer lies in the philosophical eye of the

beholder. To the diehard logical positivist, perhaps there can be no knowledge without repeated experimental test as proof. Even those at this end of the spectrum, however, might acknowledge the value of clinical experience—e.g., Newton's apple! The practitioner's more speculative approach to divining fact recognizes the inherent limits of experimentation in understanding the whole—i.e., according to the scientific method, proving that a proposition is true necessitates a detailed progression of experiments, few of which explain much of the variance in human behavior in field settings. Clinical methods, moreover, are widely respected in some scientific approaches to studying human behavior, including those used to pioneer T-group methods. Further, the *action research* tradition attempts to create situation-specific knowledge which, though not generalizable by the client, may indeed suggest patterns to the consultant who observes these phenomena across clients and time.

Transferability: Research to Practice

Some authors demand the same rigorous standards from every instrument in every psychometric situation (see Cattell & Johnson, 1986). However, this position is probably untenable and irrelevant to the reality faced by modern managers and practitioners. First, as pointed out above, many uses exist for instruments which do not necessitate measurement in the academic researcher sense. Further, economic, political, and social issues change and evolve faster than science's ability to keep up with (much less lead) all developments. Hence, industrial practice will attempt to adapt in ways which at least satisfice. Action research tools help society adapt while waiting for basic science to make fundamental breakthroughs in understanding, predicting, and controlling human behavior.

Summary and Conclusions

Approaches to ascertaining validity in various research and applied settings appear to be driven by different stakeholders toward quite different standards. All of these settings are important to society, and it would be dysfunctional for any existing approach to defining validity to overwhelm the others entirely. Nonetheless, each approach also appears to miss some of the benefits of the others. One possible reaction to this situation is for research, practice, and teaching to simply maintain their separate ways of viewing validity, though this promises to perpetuate the problems of transferability from one setting to another, thus slowing advancement. This approach is consistent with common practice in most universities. They are organized by academic disciplines, which inhibits interdisciplinary or theory/application integration while maintaining disciplinary traditions.

An alternative is for all of these standards to be collapsed into one inclusive set, constituting a higher standard toward which all activities should aspire. This would require an epic paradigm shift in academe, where the idea of "relevance" is viewed with great suspicion because it smacks of infringement upon academic freedom (i.e., that industry could thereby dictate what questions the researcher

must ask or even what findings they must report). This problem could be addressed by composing research teams of professionals from across the roles discussed in our four scenarios. Several entities seem to possess potential levers for encouraging such thrusts. One is academic accreditation organizations (e.g., American Assembly of Collegiate Schools of Business), which could recognize and reward colleges that build faculties containing both excellent theoreticians and applied experts. In addition, academic center directors and department heads might also assume leadership in encouraging programmatic research which builds from and toward both applied needs and conceptual clarification. Perhaps there is also a need for intensive, mid-career cross training beyond the doctoral level for professionals who do well in either applied or theoretical matters, but not in both.

Another alternative is that government and private funding sources increase support to institutions which view science as a servant of mankind and not simply an end in and of itself. These practitioners and researchers will willingly seek out ways to work collaboratively, striving toward the application of all the validity criteria discussed here. Many excellent research centers have roots into both academe and industry, (e.g., Center for Effective Organizations and Center for Creative Leadership). Further, just as physics moved ahead dramatically with the Manhattan Project and many sciences advanced through the Apollo Program, so OD might be advanced through a "big science" initiative from government or from industry consortia. It has been the objective of this paper to provide a reference point for constructive interaction by promoting an appreciation for a range of criteria for judging questionnaire validity. It is our belief that theory and practice, if more closely coupled, can enhance one another. This integration seems most likely to occur if we have the courage to broaden our criteria for validity.

References

- Argyris, C. (1968). Some unintended consequences of rigorous research. *Psychological Bulletin*, 7, 185-197.
- Argyris, C. (1970). *Intervention theory and method*. Reading, MA: Addison-Wesley.
- Argyris, C., Putnam, R., & Smith (1985). *Action science*. San Francisco, CA: Jossey-Bass.
- Adair, J. G., & Schachter, B. S. (1972). To cooperate or to look good? The subjects' and experimenters' perceptions of each others' intentions. *Journal of Experimental Social Psychology*, 8, 74-85.
- Bowen, D. D., & Jackson, C. N. (1985). Curing those ol' "omigod-not-another-group-class blues." *Organization Behavior Teaching Journal*, 10, 21-31.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cascio, W. F. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Reston.
- Cattell, R. B., & Johnson, R. C. (1986). *Functional psychological testing: Principles and instrument*. New York: Brunner/Mazel.

- Cook, T. D., & Perrin, B. F. (1971). The effects of suspiciousness of deception and perceived legitimacy of deception on task performance in an attitude change experiment. *Journal of Personality, 39*, 204–224.
- Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1–14.
- Freedman, R. D., & Stumpf, S. A. (1978). What can we learn from the learning style inventory? *Academy of Management Journal, 21*, 275–282.
- Freedman, R. D., & Stumpf, S. A. (1980). Learning style: Less than meets the eye. *Academy of Management Review, 23*, 445–447.
- Hackman, J. R., & Oldham, G. (1980). *Work redesign*. Reading, MA: Addison–Wesley.
- Jackson, C. N., & Manning, M. R. (in press). Evaluation fundamentals and realities. In C. N. Jackson & M. R. Manning (Eds.), *Organization development annual: Evaluating OD interventions* (Vol. 5). Alexandria, VA: American Society for Training and Development.
- Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart, and Winston.
- King, D. C., Sherwood, J. J., & Manning, M. R. (1978). OD's research base: How to expand and utilize it. In W. W. Burke (Ed.), *The cutting edge: Current theory and practice in organization development* (pp. 133–148). La Jolla, CA: University Associates.
- Kolb, D. A. (1976). *The Learning Style Inventory: Self-scoring test*. Boston, MA: McBer.
- Myers, I. B. (1962). *The Myers–Briggs Type Indicator: Manual*. Princeton, NJ: Educational Testing Service.
- Orne, M. T. (1962). On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.
- Reynolds, C. R., & Wilson, V. L. (1985). *Methodological and statistical advances in the study of individual differences*. New York: Plenum.

Biographical Note

Michael R. Manning

Department of Management
New Mexico State University
Las Cruces, NM 88003
Phone/Fax: 505–646–2532/6155
E-Mail: Mmanning@nmsuvm1.nmsu.edu

Dr. Manning is an Associate Professor in the College of Business Administration and Economics, New Mexico State University. He received his Ph.D. in organizational behavior from Purdue University and has served previously on the faculties of Case Western Reserve University and the State University of New York at Binghamton. His research interests are in organizational change and development, organizing for collective action, and individual and organizational health and well-being.

Conrad N. Jackson is Associate Professor of Management at the University of Alabama in Huntsville. He received his Ph.D. in administrative science from Purdue University and served previously on the faculties at the University of Tulsa and Western Illinois University. His research interests include factors in team success and productivity of knowledge work-

ers. His consulting clients have included Boeing, Ford Motor Company, Phillips Petroleum Company, and the US. Army Corps of Engineers.

Valaya Pathi is Associate Professor of Managerial Studies at Tiffin University. His Ph.D. is in organizational behavior from Case Western Reserve University. His research interests include choice behavior in organizations and organizational measurement.

Received: November 5, 1993

Accepted after two revisions: July 15, 1994

