Assessment as Practice: Notes on Measures, Tests, and Targets

Brigitte Jordan and Peter Putz

Drawing on data from workplace and learning studies that we have conducted over the past 20 years, we propose to rethink assessment by developing a three-part framework that puts assessments into a broader social context. The framework identifies inherent assessments as happening informally and nonverbally in all social situations; discursive assessments as occurring when members of a social group talk about what they are doing in an evaluative way; and documentary assessments as coming about when activities are evaluated according to a scheme that produces numbers and symbols. Formal, documentary assessments are ubiquitous in all arenas of modern life, from production work to corporate strategy, governmental resource allocation, and educational policy. However, they frequently have negative consequences which remain largely unexamined, in the literature as well as in the daily practice of managers and decision makers. We show that an overreliance on documentary assessments can lead to far-reaching dysfunctional effects on work practices, on corporate decision making, and on the structure and culture of an organization. In the final part of the paper, we apply our three-part assessment framework to propose a set of recommendations for managers and researchers that promise to lead to the improvement of assessment practice.

Key words: assessment, evaluation, side effects, accounting, organizational culture

There are two equally important observations that emanate from our work on learning and assessment at the Palo Alto Research Center (PARC) and the Institute for Research on Learning (IRL): 1) assessment is a normal, ubiquitous part of all social interaction; and 2) formal assessment methods as used in organizations frequently lead to undesirable results.

These observations are not of a kind. They are like dinosaurs and humans, both bipedal but otherwise occupying rather distinct life spheres. Nevertheless, we might learn something by considering them together. It may just be the case that part of the problem with assessment is precisely that the first observation has not been taken seriously. In this paper we will talk about both and attempt to draw out their importance for rethinking assessment in workplaces and schools. The central proposition here is that, in addition to the standard tests and performance measurements that are routinely administered in our institutions, there are other important forms of assessment that are not usually recognized. These can give us valuable insights and provide leverage for restructuring the way assessment systems are designed.

What emerges out of our work is a framework that complements and amplifies recent thinking around measurement practice, puts formal assessment in perspective, and recognizes it as only one piece (albeit a significant one) of the varieties of judgments about performance that play a crucial role in schools, work places, and everyday life.

Throughout the paper we use the general term “assessment” to comprehensively refer to the totality of informal and formal judgments, evaluations, measurements, tests, surveys, and metrics that play a role in productive social interaction. We start our discussion with a characterization of two kinds of assessments that are produced on the fly, as natural parts of mundane social activities by individuals and groups: inherent assessments and discursive assessments. We then contrast these with formal, standardized measurements used in organizations, which we call documentary assessments. We will show how each of these three assessment types plays a distinct role in articulating the work of individuals...
and groups on various levels. We will also show how formal, documentary assessments regularly produce dysfunctional behavioral effects because they are disconnected or in opposition to the intrinsic requirements of everyday work practices. In a final section, we will suggest implications for further research in the organizational practice of assessments and provide recommendations to managers for the improvement of assessment practice.

Assessment as Social Practice

Our aim to integrate formal assessment practices and informal assessments, which occur inevitably in the course of any mundane or professional activity, continues recent debates in the field of managerial accounting. In the 1980s critical research began to study accounting as a social and institutional practice, one that is intrinsic to, and constitutive of social relations, rather than derivative or secondary” (Miller 1994:1). This emphasis on social relationships marks a substantive shift in perspective (Hopwood and Miller 1994). It rejects any position that assumes accounting tools would produce some sort of objective representation of empirical facts while remaining otherwise neutral. It turns against what was mainstream up until the last 30 years and what is still a rather prominent perception: that accounting is merely a quest for continuous technical refinements of bookkeeping and cost-accounting schemes designed to deliver relevant information to executive decision makers, to support rational resource allocation, and to maintain institutional accountability to enhance overall efficiency of an organization (AAA 1970; Burchell et al. 1980).

The new social practice view also supersedes a purely functionalist perspective that studied the behavioral impacts of assessments on individuals (and sometimes groups), while taking the accounting scheme for granted (Hopwood 1983). These functionalist studies detached accounting processes from their organizational context, concentrating on experimental laboratory studies that revealed the cognitive and motivational effects of accounting on individuals. (For comprehensive reviews see Birnberg, Lawrenz, and Young 1983 and Sprinkle 2003.)

Under the new social paradigm, accounting becomes visible as a set of calculative practices that actively produce the social and organizational realities we live in and which are in turn constantly altered and reproduced by being enacted in everyday practice (Giddens 1984). The complex and multifaceted interplay between calculative practices and other managerial and social practices becomes the subject of a rich variety of alternative accounting approaches (for overviews see Baxter and Chua 2003; Miller 1994) which highlights the technical and transformative capacities of accounting (Boland 1981; Cooper, Hayes, and Wolf 1981; Hedberg and Jönsson 1978), its potential as rationale and language (Covaleski and Drsmith 1988; Miller and O’Leary 1987, 1994; Preston and Oakes 2001), and its ideological dimensions (Armstrong 1987; Loft 1986; Miller and O’Leary 1987; Tinker, Merino, and Neimark 1982).

The new paradigm treats assessment as a specific kind of social practice which is manifest in the construction and use of various calculi and in specific forms of interactions such as accounting talk (Ahrens 1996, 1997). However, such a perspective fails to recognize that assessments happen not only in the context of specific organizationally located accounting procedures; they also are an inevitable part of any human activity.

On the other hand, there are several social science fields that have concerned themselves with mundane assessment practices in face-to-face interaction. Scholars in kinesics, proxemics, social psychology, sociology, and later ethnography, conversation analysis, and interaction analysis, have, often in detailed, video-supported studies, developed methods and theoretical constructs for identifying the basic building blocks of human behavior (Birdwhistell 1970; Garfinkel 1967; Goffman 1959, 1963; Goodwin and Heritage 1990; Hall 1966; Jordan and Henderson 1995; Kendon 1990; Nardi and Engeström 1999; Sacks 1992; Sacks, Schegloff, and Jefferson 1974). The study of animal behavior (ethology) has also been influential in coming to understand the role of nonverbal features in interactions.

This research makes clear that social life becomes possible in the first place because of the mutual assessments that human beings continuously carry out, not only and obviously in verbal interactions, but also through nonsemantic features of language such as prosody, and through nonverbal body language including gestures, body orientation, participation frameworks, and bodily distancing. While out of awareness, the nonverbal rules of face-to-face interaction are what enable humans (and other social species) to learn from each other, relate to each other, indicate agreement or disagreement, and demonstrate group membership.

What we find, then, is an accounting literature that, though increasingly socially oriented, has focused on formal evaluation; and a social science literature that, though inherently socially focused, has been little concerned with the effects of formal assessment procedures. In this paper we will address that gap and discuss what can be gained by looking at assessments both as a privileged type of organizational practice and as a ubiquitous feature of everyday social interaction.

Methods

The data on which we draw for the framework we propose come from years of ethnographic work in a variety of settings, ranging from the United States and Europe to developing countries. They focus on a variety of topics, from formal survey processes and testing situations in schools and workplaces to detailed investigations of how people assess each other as they manage the intricate dance of sociality in everyday life.

In the early years we relied on standard anthropological participant observation coupled with in-situ question asking and documentary analysis, recording what we found in fieldnotes.
Table 1. Examples Based on Our Ethnographic Research

<table>
<thead>
<tr>
<th>Example</th>
<th>Source</th>
<th>Further References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother and children baking muffins</td>
<td>IAL data corpus</td>
<td>Shrager and Callanan 1991</td>
</tr>
<tr>
<td>Teacher checking on students</td>
<td>IAL data corpus</td>
<td></td>
</tr>
<tr>
<td>Midwife apprentice</td>
<td>Jordan’s ethnographic studies of childbirth in four cultures</td>
<td>Jordan 1993</td>
</tr>
<tr>
<td>Airlines operations room</td>
<td>IAL data corpus</td>
<td>Brun-Cottan et al. 1991; Jordan 1992</td>
</tr>
<tr>
<td>Farmworkers</td>
<td>Jordan’s fieldwork with migrant farm laborers in California in the early 1970s</td>
<td></td>
</tr>
<tr>
<td>Call center</td>
<td>IAL data corpus</td>
<td>Whalen, Whalen, and Henderson 2002</td>
</tr>
<tr>
<td>Airline personnel coding flight delays</td>
<td>IAL data corpus</td>
<td>Brun-Cottan et al. 1991; Jordan 1992</td>
</tr>
<tr>
<td>Telephone repair people</td>
<td>Fieldwork conducted by the Work Systems Design Group at Nynex between 1991 and 1996</td>
<td></td>
</tr>
</tbody>
</table>

Note: Interaction Analysis Laboratory (IAL) operated at Xerox PARC and IRL beginning in the late 1980s. Jordan and Henderson (1995) provide a detailed account of the methodology developed and applied by the IAL.

We produced written accounts and graphic sketches, kinship and organizational charts, photographs, maps, and other kinds of representations, which we reviewed with participants in an effort to capture inside (emic) as well as outside (etic) perspectives.

Since the early 1970s, we have increasingly employed audio- and then videorecording in addition to standard participant observation. This allowed us to generate data corpora that can be subjected to multiple analyses by multiple analysts for multiple purposes. Combining conventional field methods and stable, reanalyzable records has a variety of advantages discussed in detail in Jordan and Henderson (1995) and Jordan (1996).

The framework we propose here is based on our ethnographic research, including more then a thousand hours of collaborative video analysis in the Interaction Analysis Laboratories at IRL, PARC, and several universities where we have carried out research. In the analysis of videotapes we use an inductive mode of investigation where an initial noticing of “something interesting” leads to searching for similar cases, on the same tape or others, and possibly a return to the field site. If a pattern can be identified, we search for the conditions under which the pattern holds and try to come to some understanding of its distribution. With luck, there might be corroborating or—equally lucky—disconfirming information from the work of other investigators. Most of the examples we use to illustrate the theoretical constructs we propose here come from our own data corpus, supported by corroborative cases from the literature whenever possible. These examples are outlined in Table 1.

**Inherent Assessment**

We note that some form of assessment is a natural part of all socially situated activities. Informal assessments occur all around us, whenever human beings get together to accomplish any sort of activity in a collaborative, cooperative way. Whenever we take the time to look and notice, we find naturally occurring, unremarkable, and unremarked assessment activities that nevertheless are fundamental in making human collaboration possible. Even more fundamental is the fact that in-a-glance, on-the-fly mutual assessments underlie all of human sociality and, in fact, the sociality of all social species. The next five sections provide examples of these mutual assessments from everyday life.

**In Conversation**

A listener “looks puzzled”—the speaker rephrases what she just said. Both individuals have made an assessment. Such
mutual checks on understanding are continuously carried out when people talk to each other formally, as in speeches and interviews, or informally in casual conversation (Goodwin 1980; Goodwin 1986; Goodwin and Goodwin 1987; Pomerantz 1978, 1984).

**Mothers and Children**

Mothers are constantly called upon to assess their children’s acquisition of various skills—as are other family members, particularly siblings who are keen judges of younger sisters’ and brothers’ acceptability in play groups (or work groups in other cultures). A mother makes dozens of such judgments daily—when she gives the child a spoon rather than a fork with which to eat; when she does or does not let her cross the street alone, let her go to the neighbor’s house, or tell her to get her shoes tied (or not).

On a videotape, we observe the routine morning activities of a family. The tape shows us mother, four-year-old daughter, and baby Amy baking muffins for breakfast. Daughter is standing on a chair filling muffin cups, mother watching, and baby crawling up and pulling herself up to the table, looking on. Intense monitoring is going on of everybody by everybody. The mother at one point holds down the paper muffin cup with thumb and finger because, as daughter puts in heavy, sticky batter with a wooden spoon, the batter doesn’t loosen from spoon and pulls up the cup. Next time, daughter holds down the cup the same way. Now some batter is on her finger. She licks it off (which could be seen as cleansing it) with a quick sideways look at her mother. Mother lightly dips her finger into the muffin cup they’ve just been working on and licks it appreciatively, showing that it’s okay to have a lick for the taste and not just for the cleaning. At that point the daughter sticks her finger into the bowl and licks it and baby Amy who’s been watching the proceedings with great interest also gets her fingers into a muffin cup.

What we see here is a constant, ongoing assessment of activities, of what’s okay and what isn’t, of how to handle spills, of when a cup is “full enough.” The monitoring is mutual, by the mother to keep the children appropriately integrated in the activity and by the children to acquire the skills necessary to carry out the tasks at hand. And the baby is tracking it all.

**Teachers and Students**

In a classroom we see a teacher wandering over to a child she expects to have trouble with an assignment. She takes one look at what the child is doing and wanders back to her desk, having judged the child to be doing okay. Students as well make judgments about each other’s competence, an assessment that becomes apparent in how a joint problem is attacked. In contrast to externally imposed tests, these endogenously generated assessments are inherent in the social scene of ongoing classroom activities. They are generated by participants in the course of those very activities.

**Professional Judgment**

A Maya village midwife is called to the hut of a woman in labor (Jordan 1993). As soon as she arrives with her apprentice, a messenger calls her to another birth. She turns to the apprentice and says: “You stay here and take care of Doña Rosa; I’ll go and see about Doña Elvia.” An assessment has occurred. Without saying it in so many words, the senior midwife has judged her apprentice competent to conduct a birth on her own. Notably, this assessment occurs in a real-life situation under the pressures and requirements of getting real work done. There is no abstract claim to expertise here, nor any formal conveyance of authority, but rather, in the urgency of the situation the apprentice is judged to be competent. She is immediately called upon to demonstrate that competence—not abstractly, in a written examination, but in the conduct of a real birth, in an environment where she will be judged as to her ritual, manipulative, herbal, medical, and interactional competence by experts—other women who have had children and who also attend the woman in labor.

**From Novice to Expert**

On a videotape of an airline operations room we observe a supervisor and four operators, one of whom is new on the job. At one point the supervisor propels himself on his wheeled office chair from his own workstation into the midst of the four operators, with his back to the rookie. The supervisor is engaged in looking at a bank of video monitors, which are more easily scanned from that position than from his desk. After a while he asks softly, of no one in particular: “Has 464 landed?” The new operator punches something into her terminal and then answers: “They’re about to land.” The supervisor nods and continues to make annotations on a sheet of paper.

Again, an assessment has happened. What the supervisor—and the rookie’s coworkers who overhear the interchange—find out is that the new operator does not know how to get the required flight information (which the supervisor could have found out by checking his own computer or asking her directly) but also that she is in command of a much higher-level skill: to judge correctly when it is up to her to provide answers to generally posed questions. The assessment was accomplished by the supervisor in the course of, and incidental to, acquiring a useful piece of information, not something contrived for purposes of testing.

We call these embedded-in-life-activity assessments “inherent assessments” because they are a chronic feature of all human conduct in the ongoing flow of activities (Giddens 1984; Goffman 1963). They occur routinely, effortlessly, and unavoidably as part of any nonsolitary human activity where people rely on a shared sense of purpose. Thus, we find inherent assessments as part of routine work activities as well as in the normal home, family, and recreational activities of human beings. As a matter of fact, all of us make assessments of each other all the time, assuming that we each
can or cannot, will or will not, do certain sorts of things. Furthermore, we are constantly updating our assessments in the course of ongoing activities. Inherent assessments are not usually made explicit in the form of verbal utterances or mental descriptions but tend to remain in the sphere of practical consciousness.

In everyday social life as well as in work situations, inherent assessments are made in the interest and for purposes of the individual attempting to align (or misalign) with the group. As such, they constitute one of the fundamental mechanisms by which learning occurs, including the kinds of incidental learning we simply think of as normal parts of human development. Even a baby or toddler continuously assesses approval or disapproval of its actions by family members. Newcomers adjust their talk and nonverbal interactions to those of a work group they are entering. Neophytes become full members of communities of practice by quickly and unobtrusively monitoring responses and reactions of other members, and thereby gaining access to the group norms they need and want to adopt.

Assessment criteria are based on the affordances of ongoing situations rather than on underlying abstract skill requirements. It is the mutual assessment that tells people they are ready (or not) to take the next step and do the next task. Though tacit and implicit, inherent assessments are absolutely crucial for smooth, interpersonal interaction and for carrying out the work of a community of practice or any other social formation.

**Discursive Assessment**

Sometimes, in the course of an activity, participants may find a reason to make the unspoken, inherent assessment explicit. Parents may talk to their children about what they are already able to accomplish and what they will be able to do soon. A work group may begin to talk about how they are doing, how much more remains to be done, that a particular worker is lagging and why, the impact of defective parts on the speed of the assembly line, or the effect of a plane delay on activities in an airport. When assessment becomes explicit and shared within the group, we call it "discursive assessment."

What we mean to characterize here is the phenomenon of discursive assessments as they occur among coworkers who share a joint orientation based on their daily work practice. In such cases we find that discursive assessments most likely support work processes in a positive way. For example, if we listen in on "ops talk" (i.e., talk in the airline's operations room) mentioned earlier, we find it chock full of statements that indicate to insiders where they are in the complicated process of taking airplanes in and out. We might see a pilot leaning over the counter that separates the ops work area from a break room in which "ramp rats" (workers who service airplanes on the ramp), mechanics, and other personnel are taking their breaks. He says to one of the operators (but loud enough so everybody in both areas can hear): "They are changing the oil filter on number one." The operator, nodding, responds "Right," and the pilot goes on: "I asked them if this would involve a delay? They said it's gonna be close. So I thought you might like to know and the gate might like to know also...." He trails off. The operator nods repeatedly while the pilot talks. Then he says: "Great! Uh, is this a...what's the problem involved?" To which the pilot responds by explaining the technical reason for the trouble ("The bypass pin has popped out. Meaning the filter is contaminated") and what is being done about that.

The implications of this kind of assessment of the state of the work (who else needs to act? who needs to be notified? what happens if the problem can't be fixed?) are contextually clear to all those present or, if not, become a topic for further talk and negotiations that determine what needs to be done. This type of discursive assessment is not only exceedingly common but also absolutely essential for a smooth work flow.

Discursive assessments—like inherent assessments—are generated within the group to figure out, collaboratively, what state the group is in and what to do about that. The difference is that while inherent assessments rely on individual nonverbal monitoring, resulting in individual behavioral adjustment, discursive assessments make issues public, propose common standards, suggest and enforce divisions of labor, and monitor group behavior such that the work will get done. They generate information and documentation not only for internal use but also as a way to justify the group's actions, document its successes, or account for its failures externally.

Note that discursive assessments are socially mobile. They can be referred to, agreed with, or revised by people who are part of the group. They have become "social objects" that have a life within the group. Unlike inherent assessments, they have persistence, such that people can refer to them at a later time.

Though unacknowledged, the ability to talk about the ongoing activity in an evaluative way—to produce a discursive assessment—is crucial. By showing competence in group-endogenous assessment activities, members, and particularly new members, demonstrate their proficiency and gain acceptance. Discursive assessments are exceedingly common in work situations where a continuous, real-time evaluation of a particular situation or procedure is imperative to update all parties involved about the current state of their world. Routine work is filled with huge numbers of these.

Discursive assessment, the reflective talk generated by a group of people engaged in a particular activity about that activity, is an important vehicle for learning and innovation. As activity theorists have pointed out, group-internal discursive assessment functions to facilitate reflection and thereby plays a significant role in learning (Nardi 1996; Wertsch 1981). It creates a shared understanding of individual roles and responsibilities and thereby works out a division of labor. At the same time, these informal assessments create
a public verbal representation of the capabilities, resources, and issues for a group that enable them to consider implications of the current state, as they understand it, for behaving more effectively. Like inherent assessments, they are often produced on the fly, effortlessly, without official training, as part of the ongoing activities of schools, workplaces, and life spaces. This can happen in the hallway or the coffee shop as readily as during dedicated work activity.

**Documentary Assessment**

The third type of assessment, “documentary assessment,” is the one we are all most familiar with. It involves externally mandated, stable symbolic representations of evaluations in the form of tests, surveys, check lists, plans, targets and similar instruments.

Documentary assessment occurs when an enduring record of some kind is produced, a set of marks on a piece of paper (or on a computer), that is reflective and evaluative of some activity. In inherent and discursive assessment, no independent record gets created; the assessment just gets folded back into the activity, molds the activity, but is not externally available. In documentary assessment, on the other hand, we see an extension of mobility beyond the group. The documentary assessment now has the properties of “immutable mobiles” (Latour 1986); that is to say, once constructed, the content becomes fixed while the assessment itself becomes mobile as a document within a larger socioeconomic system. Immutability and mobility are the features that allow distant parties to grasp what is going on in remote places and to intervene there.

By virtue of the shared symbol system they employ (consisting of numbers or standard evaluative statements) documentary assessments become linking pins between diverse stakeholders involved in corporate and public affairs. In practice, assessments become “boundary objects,” objects that are plastic enough to adapt to local needs and the constraints of several parties, yet robust enough to maintain a common identity across sites (Star and Griesemer 1989).

Documentary assessment occurs primarily to evaluate the extent to which preestablished performance targets have been achieved and to establish cross-group comparability. In schools, a teacher’s entries in her grade book or students’ scores on standardized tests are less designed to facilitate students’ learning than to satisfy the requirements of ranking this student among others (the teacher’s job) or ranking this school among schools in the district and the state (the school superintendent’s job). Such ranking is carried out to respond to the interests of political constituencies and not primarily to the interests of the students engaged in the business of learning.

Comparison is the most salient operation in this context. For the sake of comparison, the great variety of real-world phenomena has to be reduced. If we take an ongoing, lived experience, and if we then produce a written representation of that (fill out a form, answer a questionnaire), certain consequences ensue. We are no longer operating in the lived world of experience but within a symbol system. This is known as translation of an experiential analog system to a symbolic digital system, and this translation always involves loss of contextual information. The loss can be significant. If the teacher gives the student a “C” or the supervisor assigns a “satisfactory” to a worker’s performance, this makes that particular entry comparable to others, but shears it of its local context.

However, the very loss of contextual information buys us two important features: data comparability and expanded distribution. Stripping the context is precisely what makes comparison possible without regard to individual circumstances and this is why documentary assessment is always called for when “objectivity” is an issue. The intention is to move evaluation from the hands of interested parties to some sort of formal, objective procedure. Documentary assessment requires instruments that are insensitive to differences in ethnicity, gender, social status, work and learning environment, personal histories and relationships, and all other potential sources of bias. How is that possible? One distinction between naturally occurring (inherent and discursive) assessments and formal documentary assessments is that the former include an affective component while the latter specifically exclude that, i.e. participants do not share the same Lebenswelt (life circumstances). Expertise, as judged by endogenous assessment, is based on coexperience, while documentary assessment tends to be based on externally relevant standards.

In a deeper sense, the choice of questions to be asked and variables to be measured by these assessment instruments makes certain aspects of an organization’s activities visible and important and hides others. The very structure of the instruments and the manner of their administration can be seen as indicative of an applied theory about how a particular organization sees itself working. In this sense, documentary assessments are manifestations of a management rationale, providing a complex set of vocabularies and meanings that are linked to the educational, political, and economic concepts the organization is devoted to (Miller 1994).

From the local vantage point, however, documentary assessments are often perceived as imposed from the outside, carried out in the interest of superordinate stakeholders to further interests that do not directly overlap with those of the group being assessed. In fact, it is the shift from a group-internal locus of control to an external locus of control that characterizes the boundary between discursive and documentary assessment.

In the workplace, the mutual adjustments made possible through inherent assessment and the shared understanding generated by discursive assessment are now joined by the requirements of accounting for success or failure externally (i.e., measuring performance according to bureaucratic standards). We will see that the consequences of externally mandated assessments for what happens in workplaces and school systems can be highly problematic.
Table 2. Characteristics of Different Assessment Types

<table>
<thead>
<tr>
<th></th>
<th>Inherent Assessment</th>
<th>Discursive Assessment</th>
<th>Documentary Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary stakeholder</td>
<td>Individual</td>
<td>Team/social group</td>
<td>Corporate entities</td>
</tr>
<tr>
<td>Purpose</td>
<td>Adjust individual conduct</td>
<td>Align group-internal views</td>
<td>Coordinate and control organizations</td>
</tr>
<tr>
<td>Interaction</td>
<td>Physical copresence</td>
<td>Physical and virtual copresence</td>
<td>Distance</td>
</tr>
<tr>
<td>Mode of coordination</td>
<td>Nonverbal monitoring</td>
<td>Conversation</td>
<td>Record production</td>
</tr>
<tr>
<td>Locus of control</td>
<td>Individual</td>
<td>Group-internal</td>
<td>External</td>
</tr>
<tr>
<td>Knowledge produced</td>
<td>Tacit</td>
<td>Implicit/explicit</td>
<td>Explicit</td>
</tr>
<tr>
<td>Duration</td>
<td>Ephemeral</td>
<td>Short lived</td>
<td>Enduring</td>
</tr>
<tr>
<td>Emotional tone</td>
<td>Affective</td>
<td>Affective</td>
<td>Objective</td>
</tr>
</tbody>
</table>

From Inherent to Discursive to Documentary Assessment: An Extended Example

Let us look at a real-world situation where assessment activities are rampant, as they are in all work situations.

A group of farmworkers is harvesting onions. Each worker goes down a row, pulls up a bundle of onions by the withered greens, holds the bunch over a burlap bag, and cuts the greens with a pair of clippers so that the onions fall into the sack. As the sack gets full, the worker carries it to the end of the row, where it will be counted by the foreman and picked up by a truck.

Now, at any time during the work, be it for purposes of coordinating a break with a friend or for seeing who’s ahead or who’s not keeping up and might need help, a sweeping glance over the field is all that is necessary for an assessment. If they all started at the edge of the field, their very position shows if they are “ahead” or “behind.” In addition, the quantity of bags at the end of rows provides an immediate measure (in volume, even if not counted) of such things as how far along they are toward the next pickup. These assessments are inherent to the activity; you couldn’t be a team member without making them and they provide valuable resources for the sequencing of subtasks, for energy expenditure, for judgments about a person’s strength, state of health, needs for money, and the like.

At some point, assessment becomes discursive. One of the workers suggests that if they each did seven more sacks they would be done with the field. At that point, people begin to count how much they’ve done already and estimate how many more onions are out there to be picked. The assessment becomes a topic for discourse in the community; judgments are made, questioned, negotiated, but note that this is done entirely for purposes of the group.

Now the foreman comes and writes down numbers for each crew member. At this point, a set of marks is created that shows characteristics of its own—the numbers can be added; an average can be computed; who has picked the most and the amount of money due each worker can be determined. At the same time, looking only at the numbers without having available the field and the situation of the workers loses information. For example, anybody looking at the field can see that one border of the field has much smaller onions than the rest. So the fact that the worker who had those rows produced fewer bags is not to be seen as lower productivity. We also lose information on how tightly filled the bags are or who is helping whom. We call the foreman’s record keeping a documentary assessment. It provides a stable rather than evanescent record of activities, in contrast to inherent and discursive assessment where the assessment molds the activity, but is not externally available.

In the evening, when the foreman compiles the numbers for his crews, he doesn’t care how “hard” or under what circumstances people worked, he only cares about what numbers he has to deliver to the big boss who will use them to pay the workers at the end of the week. He could single out particular workers as “bad apples,” but in this particular case, since these work groups are comparatively autonomous, held together by ties of blood and fictive kinship, he won’t bother to do that (though that is definitely a dreaded possibility in many work situations).

To summarize the characteristic features of the three types of assessment, consider Table 2.

The Limitations of Documentary Assessments

Unanticipated, unwelcome consequences of overreliance on formal documentary measurements are well known
to workers and managers alike. Though they have not been investigated systematically by accounting research, students from disciplines such as the anthropology of work and computer-supported cooperative work (CSWCW) know that the introduction of formal evaluation schemes often generates unanticipated “work-arounds” that undermine the intent of the assessment. We see the potential negative consequences of documentary assessments as falling into three categories: 1) manipulating the numbers; 2) changing work practices; and 3) modifying organizational structure, climate, and culture.

Sometimes only one or two of these occur. At other times, however, an appalling blend of all three can be observed. For example, when call center operators forward a troublesome call to cut it short so they can make their targets, this simultaneously changes call statistics, affects work practices, and influences resource allocation and customer satisfaction.

_Fixing the numbers_ is a coping strategy that tends to emerge when the expectations communicated by the assessment instrument cannot be met. In that case, the numbers entered into the official record may actually be different from those an independent observer would enter. Thus, airline personnel engaged in delay coding may enter a delay into a different category than the one into which it belongs if the true category is dangerously close to reaching the allowable limit, and telephone repair people may enter a “customer not at home” code for justifying service calls that were aborted under the pressure of an overwhelming workload. In general, when performance data get used to punish and reward, a strong motivation is generated to manipulate the numbers (Kohn 1999). The game becomes one of making the numbers look good rather than improving the learning or work process.

It is obvious that such strategies produce flawed performance data, which subsequently may lead to wrong managerial decisions. This can produce grave overestimation of productivity and other crucial indicators of organizational well-being. Given the central role numerical data, trends, and projections play in managerial decision making, this is a serious issue.

_Changes in how work is actually carried out_ often occur as people adjust to the new regulations. While intended to increase productivity and quality in business and education, what often occurs is the opposite. This is particularly likely when allocation for education funds depend on performance criteria or where evaluation requirements have been imposed from the outside, as in certification and accreditation procedures. In that case, teachers and school administrators, workers and line managers may collaborate to make things look right, with the silent agreement of top decision makers.

For example, when Quality Standard norms were introduced in the Mexican auto parts industry, normal work practices were significantly affected by all kinds of cosmetic—yet time- and cost-intensive—activities, especially before impending audits. Workplaces were cleaned and sanitized, all evidence of nonstandard practices was removed, and employees spent considerable time being coached to answer auditors’ questions appropriately (Bueno Castellanos 2001).

Similarly, in the educational literature, evidence abounds that teachers “teach to the test,” developing strategies for improving the test results for their classes: going over exams from recent years, emphasizing potential exam questions in their teaching, coaching students in strategies for taking examinations (Popham 2001). Spending teaching time on such issues teaches students to know how to take tests rather than solid knowledge of the subject matter.

In cases of this sort, changes occur in how the work is actually carried out after the introduction of formal measurement systems. But these adaptations make the process as a whole more cumbersome and more expensive since the time and effort spent in making these adjustments do not actually further the work that has to be accomplished. The persons or groups assessed change their practice in such a way that their performance, as it appears on the established indicators, is optimized, to the detriment of other objectives or efficiency criteria outside of the scope of documentary assessments.

Sometimes the introduction of a documentary assessment also triggers undesired changes in organizational climate and culture. In some situations insidious changes occur that, appearing gradually, may remain “invisible in plain sight.” For one thing, there is likely to be an increase in personnel that is concerned with collecting and analyzing relevant statistics. Administration of formal measurement requires a new layer of bureaucratic experts and supervisors and a new set of specific procedures, not so much concerned with manufacturing auto parts, teaching students, or providing customer service, but with providing accounting evidence.

More subtle, however, may be changes that occur in the culture of work. When assessment data are compared across individuals or across plants, new levels of accountability emerge. Employees begin to look out for themselves to avoid retribution rather than focusing on shared goals. There arises a temptation to fix the numbers, to focus on short-term success, and to deny responsibility for failure. As a consequence one may see mistrust, competitiveness, passing-the-buck behavior, and what is known in the workplace as a “cover-your-ass” (CYA) attitude. Some organizational theorists (e.g., Gittell 2000) have argued that quantitative performance measurements inevitably generate some level of dysfunctional behavior, since they tend to operate with a relatively low level of trust. Indeed, they are virtually guaranteed to produce low trust by setting up conflicts over who is responsible for problems. Stifling creativity and out-of-the-box thinking in favor of conformity and towing the line deprives the organization of otherwise available learning and innovation opportunities.

Baker (1994), writing in the _Fifth Discipline Fieldbook_, provides a telling example of how the spirit of the Quality Movement was undermined by overzealous measurement at Ford Motor Company. Deming’s original idea was that the work group would decide which of the hundreds of possible measurements of their work process was most valuable to give
them for continuous feedback about their performance. But when Ford management asked their plants for more figures, a number of interesting things began to happen. Indexes were developed and measurements were compared across plants. A whole new crew of bureaucratic intermediaries emerged to perform operations such as computing averages and comparing highs and lows across plants and across quarters. This led to new levels of accountability—no longer were the assessment data used only to improve the group’s performance, but team leaders now had to explain why their figures “weren’t up to snuff” and what plans they had made to improve them. The game in each plant changed from making improvements to making the plant look good in relation to its neighbors.

Gittell (2000) found similar effects in the airline industry. She studied coordination mechanisms for optimizing flight-departure processes—a highly dependent collaboration among gate agents, baggage agents, caterers, fuelers, cabin cleaners, flight attendants, pilots, and customers. In its effort to minimize expensive delays, American Airlines increased its supervisory span and focused on accountability. It established a report system that forces the manager on duty to determine which function has caused the delay. There were some unintended side effects. Employees looked out for themselves to avoid recrimination rather than focusing on shared goals of on-time performance, accurate baggage handling, and satisfied customers. The most frequent problem was communication breakdown, but since the company had no code for that, managers tended to assign such errors to the last group off the plane. Unproductive debates, reports, and meetings to determine cause of delays occurred more frequently, undermining the intent of the new system.

Drawing attention to the problematic side of documentary assessments is not to be misunderstood as an argument against the use of documentary assessments per se. Documentary assessments are fruitfully and appropriately used when the objective is comparison of large organizational units or subpopulations. In that case, stripping away the context, the why’s and how’s of a situation, and the confusion of different circumstances, is precisely what is necessary to make cross-unit comparison possible. In some managerial decision-making situations this is desirable. In others, the price may be too high.

Implications for Researchers and Managers

Distinguishing inherent, discursive, and documentary assessments and acknowledging their significance for the coordination and control of modern organizations will—we hope—help both researchers and managers to broaden their perspective beyond the dominant focus on traditional, yet often dysfunctional, documentary assessments. In the following two sections we will outline how this three-way assessment framework could generate a novel research agenda as well as an approach for managers to create an environment supportive of all forms of assessments, meeting the different needs of individuals, groups, and the entire organization alike.

Implications for Research

Our framework of inherent, discursive, and documentary assessment has its roots in detailed observations of work practices in situ. Future research needs to extend such investigations systematically and carefully to deepen our understanding of these analytically distinct yet practically often intermingled evaluative activities.

Given their impact on productivity and worker morale, the detailed study of the work-arounds, short-cuts, and exceptions that occur in all work and learning environments would be particularly fruitful. We need to understand these phenomena much better to accurately describe the impact documentary assessments have on work and teaching practice and to understand how they affect the validity of the data on which managers base their decisions.

The premier research methods for this kind of investigation are ethnographic methods, in particular participant observation aided by videotape analysis, mixed with real-time question-asking and documentary analysis (Blomberg et al. 1993; Eisenhardt 1989; Jordan 1996, 1997; Jordan and Henderson 1995). Given these methodological choices, research on inherent assessments must be grounded in observing individuals pursuing their everyday work. One might want to pay special attention to interactions where instances of adjustments and alignments with coworkers and customers occur. “Hitches” in interactions and how people fix them are good opportunities to watch inherent assessments in action. Following newcomers to find out how they manage to fit into their position, how they learn to align their own activities with the activities of others to fulfill the requirements of their tasks, will also prove fruitful.

A whole new research agenda opens up in the investigation of inherent assessment in geographically distributed teams. At this point we know very little about how new communication technologies affect inherent assessment in distance collaboration. Preliminary studies have shown that even video-supported interactions between remote teams suffer from the effects of technology-generated delays in transmission, affecting the establishment of trust and judgments of competence (Ruhleder and Jordan 2001). To what extent these technologies undermine inherent assessment (and, for that matter, discursive assessment) is unknown at this time.

Discursive assessments can be found in all kinds of formal and informal get-togethers. Work groups with a high level of internal control, such as research teams, groups of consultants, and semiautonomous work groups, might already dedicate parts of their regular meetings to internal reflection and evaluation. These are good places to look. Whenever we have kept track of employees’ talk during lunch, we found that often a substantial part of lunchroom conversations has to do with work issues. Here is where assessment becomes discursive and where it is possible to learn how the common views get built that provide the basis for coordinated efforts later in the workday. In general, places where people can get
together casually such as hallways, water coolers, and coffee machines are good venues for studying group-endogenous assessment practices.

It is also worth considering that different kinds of work provide different kinds of homes for discursive assessments, depending on the temporal structure of the work. For example, in an airlines operations room we studied, the work revolved around nine “complexes” per day. During each complex, a group of planes comes in, exchanges passengers, crews, and baggage, gets serviced, and moves out again. The sequence of complexes establishes the rhythm of the work. We found that discursive assessments tended to occur with increased frequency during the downtime between the complexes. We have also observed such rhythms in software development teams producing a new program version, construction crews moving from site to site, and teachers meeting during breaks between their classes and at certain times throughout a semester. Most of these cycles include slow-down periods, where joint group reflection takes place with increased frequency. In work that is relatively uniform and linear, like the flow on classic assembly lines or the steady stream of incoming calls in call centers, discursive assessments are more likely to be conducted in the form of conversations and chats in locker rooms and cafeterias, parking lots, and among smokers gathering outside the back door.

Apart from studying the work processes and practices of the people who collect and analyze documentary data, a novel research direction may be to look at “accounting talk” (Ahrens 1996, 1997) in all its forms, how it unfolds in informal and formal meetings, and how it affects a wide range of decisions, perceptions, and objectives (cf. Hanks 1991). Accounting talk is concerned with data produced by documentary methods, that is to say, with numbers, statistics, and trends. It takes place in public debates within school districts, parent-teacher discussions, and school board meetings. In business organizations, you may find it in quarterly ops reviews, budget meetings, planning sessions, and employee performance appraisals, where it guides and controls everyday work processes of front-line workers and managers alike.

To summarize: Research most urgently needs to turn its attention to the task of gathering systematic empirical data on assessment practices as they occur naturally in various organizational contexts. Detailed ethnographic studies that could elucidate the “talking and doing” of assessment are much needed to deepen and fully develop a social practice view on assessment and evaluation.

**Implications for Management**

The three-part framework of inherent, discursive, and documentary assessments we are proposing delivers two main messages to managers:

1. don’t focus on documentary assessments alone; nurture and support inherent and discursive assessments; and
2. if documentary assessments are carried out, be critical and reflective about their purpose, their methods and their effects.

This will require some measure of expertise to identify positive as well as negative work-arounds and other consequences of documentary assessments by managers themselves or through hiring social science experts in ethnography who can undertake such investigations. Interestingly, a number of large corporations, including Xerox, Motorola, and Intel, now have applied anthropologists, sociologists, and psychologists on their staff who specialize in such analyses.

One might think that inherent assessments are beyond management’s sphere of influence since they are accomplished by individuals on the fly, seamlessly integrated in their work activities, and, therefore, often unrecognized. This is certainly true. However, organizations can create physical and social environments that make it easier for employees to acquire the tacit knowledge that allows them to function as competent members of their work group and of the entire organization. Given how important inherent assessments are for facilitating change processes and for socializing novices into functioning communities of practice, companies might want to think about the design of work environments that facilitate unproblematic alignment. For example, a study conducted in a big customer call center revealed that lower cubicle walls supported individuals in their inherent assessments by allowing easy monitoring of and adjustment to experienced workers’ verbal and nonverbal practices (Whalen and Whalen 1996).

In general, it seems advisable to create work processes that allow easy exchange among peers and to nurture informal exchange between coworkers by providing places where people can get together. Isolating workers, and particularly novices, in solitary cubicles is clearly not conducive to either inherent or discursive assessment.

**Discursive assessments** are a valuable source for supervisors and senior managers to learn “what really goes on” within a work group, a type of information that is not likely to be captured in any kind of documentary assessment. Managers need to develop a sense of where and when group-endogenous evaluations happen and how they can listen to them. We mean this not in the sense of hidden overhearing and spying, but in the sense of building an open communication framework that encourages employees to make managers part of their communicating circle. Discursive assessments can be supported actively by providing space and occasions for all kinds of informal meetings and by creating work processes and organizational structures that encourage work groups and communities of practice to engage in evaluations. The delegation of responsibility for the design of work processes and for the internal division of labor to work groups is one way to encourage discursive, group-endogenous assessments structurally.

**Documentary assessments**, as we have shown, often do not deliver what their creators had in mind. Paying attention to the work-arounds they generate in many cases can be turned
into a source of creative improvements that fix emerging shortcomings in a way that increases the efficiency of a production processes in the long run (Herbsleb and Grinter 1998).

What is critical for managers is to find out how documentary assessments actually influence local work processes and accounting practices—every time a new measurement procedure gets designed and then applied. Awareness of complementary assessment methods and their specific affordances, drawbacks, and unintended consequences could be a step toward collaborative codesign of new organizational structures within which endogenous as well as exogenous, informal as well as formal, evaluation criteria can flourish.

The Social Structuring of Assessment

New forms of assessment are going to enable and, indeed, require a restructuring of social relationships, including some mitigation of the boundaries between the interests of those being assessed and those utilizing those assessments. This is as true of the workplace as of schools. For both venues that means a move from a hierarchical social system in the direction of a collaborative one, where workers have a major stake in the assessment of their work, and where students, with the advice and support of teachers and parents, are the owners of the learning enterprise.

In workplaces this would legitimize the development of internal standards while encouraging peer learning and team efforts. In schools, rethinking our assessment procedures with this framework will change the social structure of the classroom. It will enable more effective learning by legitimizing collaborative activities in the classroom and creating environments that encourage thinking to occur.

The question that remains to be answered is: how can we negotiate the trade-off between the requirements of institutions that need comparable numbers and objective measurements for their purposes and the learners and teachers in schools and workplaces who need the endogenous assessment derived from the activity to be mastered.

Notes

1However, as groups can become rigid and pathological due to particular internal dynamics or external pressures, their assessments can turn dysfunctional. In some work situations there may exist a shifting boundary between endogenously produced discursive assessments and more or less formal post-event “review sessions”. Externally mandated, the latter tend to reflect the interests of larger organizational units rather than furthering the in-the-moment activities of the group. In such situations groupthink phenomena (Janis 1972) are more likely to emerge.

References Cited

Ahrens, Thomas

American Accounting Association (AAA)

Armstrong, Peter

Baker, Edward

Baxter, Jane, and Wai Fong Chua

Birdwhistell, Ray L.

Birnberg, Jacob G., Turopolce Lawrenz, and Mark S. Young
1983 The Organizational Context of Accounting. Accounting, Organizations and Society 8:111-129.

Blomberg, Jeannette, Jean Giacomi, Andrea Mosher, and Pat Swenton-Wall

Boland, Jr., Richard J.

Brun-Cottan, Françoise, Kathryn Forbes, Charles Goodwin, Marjorie Harness Goodwin, Brigitte Jordan, Lucy A. Suchman, and Randall Trigg

Bueno Castellanos, Carmen

Burchell, Stuart, Colin Clubb, Anthony G. Hopwood, John Hughes, and Janine Nahapiet

Cooper, David J., David Hayes, and Frank Wolf

Covaleski, Mark A., and Mark W. Dirsmith


Hedberg, Bo, and Sten Jönsson 1978 Designing Semi-Confusing Information Systems for Organizations in Change Environments. Accounting, Organizations and Society 3:47-64.


Pomerantz, Anita


Popham, W. James

Preston, Alistair, and Leslie Oakes

Ruhleder, Karen, and Brigitte Jordan

Sacks, Harvey

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson

Shrager, Jeff, and Maureen Callanan

Sprinkle, Geoffrey B.

Star, Susan Leigh, and James R. Griesemer

Tinker, Anthony M., Barbara D. Merino, and Marilyn Dale Neimark

Wertsch, James V., ed.

Whalen, Jack, and Marilyn Whalen

Whalen, Jack, Marilyn Whalen, and Kathryn Henderson