A personal, retrospective view of ecological validity

AARON V. CICOUREL

Abstract

The notion of ecological validity has been an abiding topic either implicitly or explicitly during my entire academic career. Ecological validity focuses on how we seek to convince others of the viability and authenticity of our claims and can be understood by our use of primary and secondary data sources such as official statistics, demographic distributions, sample surveys, structured interviews, open-ended or unstructured interviews, and recorded discourse during social interaction. Ecological validity, however, can only be approximated in the social and behavioral sciences. The key issue is the extent to which data are congruent with systematic time samples of events and activities within local institutional or organizational settings. For example, direct observation of and participation in the everyday activities or practices of human actors during their daily life experiences by the use of, when possible, audio or video recordings. Systematic behavioral sampling by biological (behavioral ecological) studies of nonhuman animals (Altmann 1974; Krebs and Davies 1997) can serve as a guide for social scientists. This paper selectively discusses three previous research projects that attempted to approximate ecological validity.

1. Introduction

The problem of ecological validity was first called to my attention by a former teacher, W. S. Robinson (1950), in classes at UCLA on social statistics. Robinson had written about the validity issue by reference to the dangers of assuming that census tracts were representative of individual behavior in locally organized urban settings. The creation of census tracts was often motivated by local, regional, and national social, economic and political conditions rather than the concerns of social science research.

During the years 1958–1960 while teaching at Northwestern University, I had the good fortune of meeting and jointly teaching graduate

1860–7330/07/0027–0735
Online 1860–7349

Text & Talk 27–5/6 (2007), pp. 735–752
DOI 10.1515/TEXT.2007.033

students with Donald T. Campbell. Campbell introduced me to work on ecological validity by one his former teachers in psychology at Berkeley, Egon Brunswik (1956). Brunswik's view of ecological validity examined the relationship between a psychologically defined distal variable and proximal effects (including the relationship between behavior in controlled experimental settings and approximations outside of the laboratory). Subsequent work on the notion can be found in Bronfenbrenner (1977), Cicourel (1996), Cole et al. (1982), and Neisser (1982).

One anonymous reviewer suggested the notion of ecological validity should be linked to other types of validity and specifically mentioned 'construct validity', also developed by psychologists. The notion of construct validity refers to the problem of what data indicate (in contrast to agreement about, say, scoring reliability on what the data are). The problem of validity is the problem of what the data *indicate*. In other words, have we measured what we have undertaken to measure? For example, what if an industrial psychologist devises a test for selecting men or women who will be able to do well on some job or task? The validity of the test becomes the correlation between test scores on some instrument and, say, the quality of job performance. Note that 'performance' can mean observing the amount of 'output' on a task, such as accurately processing insurance claims within some period of time, and the like.

Validity in the non-experimental social sciences refers to the extent to which complex organizational activities represented by aggregated data from public and private sources and demographic and sample surveys can be linked to the collection, integration, and assessment of temporal samples of observable (and when possible recordable) activities in daily life settings.

Fragments of discourse materials always are shaped and constrained by the larger organizational settings in which they emerge and simultaneously influenced by cognitive/emotional processes despite the convenience of only focusing on extracted fragments independently of the organizational and cognitive/emotional complexity of daily life settings.

Social interaction and discourse materials are sometimes portrayed by applied linguists and/or students of conversation as if they have a life of their own (Schegloff 1987). This latter view, however, ignores the fact that situated social interaction is always embedded in daily life socio-cultural and cognitive/emotional processes that constrain and shape discourse. The challenge remains how daily life activities simultaneously constrain and shape more complex organizational structures (Cicourel 1992, 2006).

The validity of inferences from the study of observations and recordings of social interaction and discourse presupposes complex forms of

social organization such as kinship systems, and a variety of bureaucratic and related forms of organization.

2. Trying to approximate ecological validity

In several sections below, I discuss three projects whose field research was completed by the summer of 1970. Three additional studies (not discussed below) in the years 1958–1970 also addressed elements of ecological validity. For example, a study of a high school counseling system and the selective way decisions were reached by counselors seeking to place students in a college or university regardless of their test scores and classroom grades (Cicourel and Kitsuse 1963), a study of Argentine fertility (Cicourel 1974), and studies of American and British sign language (Cicourel 1976, 1977). A different type of ecological validity is discussed briefly in the conclusion below.

The training I received in departments of anthropology, linguistics, psychology, and sociology discussed reliability but often avoided demonstrating the more illusive problem of validity. The many statistics classes I attended in mathematics, psychology, sociology were not helpful in explaining validity issues involved in linking experimental and survey/demographic studies to patterns of behavior and social interaction in daily life settings.

A few teachers suggested that systematic ethnographic field research was a way to enhance validity. I soon realized, however, that field research had its own drawbacks because of selectivity issues always associated with labor-intensive case studies. In addition, there was another problem: the long-standing criticism that field research lacks data that can create generalizations. Although the development of the cross-cultural area files appeared promising as a way of aggregating ethnographic studies, many qualitatively sensitive issues of daily life seemed to be harshly compressed by the coding process involved. At the time, I was unaware of how field research could be enhanced by incorporating behavioral sampling methods and systematic observational research on non-human animals conducted by behavioral ecologists in biology (Altmann 1974; Krebs and Davies 1997).

3. Pursuing ecological validity

The study of ecological validity requires conceptual and empirical clarification of those substantive elements of local and larger institutional environments said to be accessible and inaccessible for study. I will not

discuss experimentally designed studies, but see Cicourel (1996). The ideal types described below are intended to be suggestive, not definitive of the use of different social science research methods.

One group relies on small samples of unstructured or semi-structured interviews without or limited ethnographic field research. The focus of attention is on the substantive content of interviews that often address a current social problem, for example, delinquent or criminal behavior, ethnic or race relations, health issues, voting behavior, fertility behavior, and occupational careers.

A second group relies on recorded exchanges that are not always observed directly and where little or no reported ethnography is presented. A subset of this category uses videotapes and may focus on the use of gestures and body movements. The primary concern is with (a) turn-taking coordination of discourse activities that help structure social interaction, (b) descriptive accounts of the social action therein, often requiring indirect or unstated references to relationships among speakers that presupposes the analyst's taken for granted common sense knowledge of local and larger socio-cultural conditions, (c) the study of preference rules for person reference, and (d) a general concern for aspects of universal systematics of language use.

Sampling across ethnographic or organizational settings is seldom an explicit topic or activity for the narrow focus on discourse. Nor is the problem of how access to sources of data are achieved and the researcher's relationship to the subjects recorded. For example, how might these latter conditions have contributed to being able to record interaction and explain particular forms of language use and gestures? Validity is viewed as an inherent quality of 'naturally occurring' and formal and informal 'institutional' talk.

A smaller group believes that direct observation and participation in a group's or community's everyday existence can reveal palpable patterns of social interaction, language use, and socially organized and constrained activities that are constitutive of all human groups. The hope is that systematic ethnographic grounding of social interaction and language use can reveal essential elements of human social organization that are normally inaccessible unless there is direct access to, and observation of, expected and unexpected, routine daily life activities. Elicited information about previously observed events or narrative accounts is believed to provide additional useful evidence for seeking generalizations.

A fourth and much larger group of practitioners uses large sample surveys, and/or large corpora of secondary information gathered by private and public agencies, a use of statistical (zero-order or multivariate) correlations, and often a disdain for ethnographic field research. The

correlations used for generating hypotheses are always suggestive, and can be used to produce generalizations. Testing hypotheses with such aggregate data remains illusive because no direct observation and/or recording of daily life activities are deemed necessary. The claim is that high reliability can be achieved despite an avoidance of what constitutes validity in the use of aggregate data.

4. Ethnographic research

The continual challenge of field research is how can direct observation and participation enhance our general understanding of daily life when it is almost by necessity confined to particular case studies? Although critics claim that field research cannot produce reliable generalizations, the depth, breadth, and convincing quality of direct observations of everyday life problem-solving or decision-making often contrast sharply with the abstract results based on sample surveys, demographic sources, and official private and public statistical reports.

The creation within anthropology of the cross-cultural area files was hoped to be a resource for alleviating the problem of generalization, but these sources remain somewhat controversial because of the considerable variability among anthropologists in the way they have conducted and reported their research (Murdock 1967, 1975). The initial classification decisions by field researchers and the coding procedures employed for creating uniformity within the files can obscure the original field research conditions. Few anthropologists take seriously the study of language use in their different field settings; few use linguistic anthropological methods that date back to Franz Boas (1938) and especially Edward Sapir (Mandelbaum 1949).

One advantage of participant observation case studies is that they accentuate the relative importance of variable and patterned semantic meanings and pragmatic effects that can be observed in behavioral performances and ongoing discourse. Unexpected and routine events require that researchers be able to observe (and when possible record) a community for long periods of time and become an identifiable agent within local settings.

Ecological validity appears to be high when we can identify locally created and sanctioned participant frameworks of agency and authority. Individual and collective memories, however, cannot be assumed to be self-evident (see Neisser 1982; Loftus 1979) in field research nor in the use of other social science methods. Memory changes remain a pervasive issue in all research on humans.

Bureaucratic and tribal rituals, and situated, routine everyday practices are associated with different types of speech registers that are often domain-specific despite the ability of some participants to move from one (say) everyday setting to another (say) technical domain that are presumed to facilitate understanding and collaboration or conflict.

5. Ecological validity in survey research: contrasting views of retired persons

The concern with the interplay of theoretical, substantive, and methodological issues began when I was the research assistant responsible for developing a questionnaire for a 5000-respondent, nationwide sample survey of everyday activities after retirement. I was hired because of my complaints to a member of my dissertation committee that sample surveys were often not constructed by explicit attempts to operationalize theoretical concepts using specific questions.

My charge was to develop a theoretically relevant questionnaire for the project. I examined theoretical work on social stratification and labor force participation and then derived a first draft of a questionnaire. After a pretest of the questionnaire in Yonkers, New York, and an area in Connecticut, I joined the two principal investigators (PIs) interviewing part of the nationwide sample in Rochester, New York. When we returned to the university, we met to discuss discrepancies between what was expected from the pre-test questions and what was obtained.

The two PIs, relying on their prior experiences with pre-tests, immediately began changing the wording of the questions. I took notes during our meeting and suggested that it was important for me to go back to the theoretical concepts to discern what should be changed in the theory in order to be consistent with the reasons given for the revised questions. I was told there was no time for such changes because of practical circumstances; it was necessary to create a new questionnaire in order to meet a deadline set by the granting agency.

The PIs became irritated with my preoccupation to clarify the original theoretical concepts from which I created the survey questionnaire used in the pre-test. My objections almost cost me my position at a time when my wife was expecting our first child and would soon stop working, and I was trying to complete my dissertation. I continued as the research assistant but no longer challenged the two PIs. I will not cite appropriate references because I remain grateful to the principal investigators. They continued to support my much needed research assistantship despite our disagreements.

I sought an ethnographic study of retired persons in the local community for my dissertation to clarify doubts I had developed from my research assistant experiences with the sample survey.

I devoted considerable time gaining daily access to a 'Senior Citizens Club'. I visited the club regularly, eating lunch with members, attending parties and dancing with my consultants. I also volunteered to organize a men's group for outings on Saturdays. The contacts acquired enabled me to interview individuals in their homes under informal conditions. The field research precluded a large sample of subjects.

Labor-intensive field methods, therefore, enabled me to obtain a sense of the daily lives of retired persons that contrasted with the aggregated results of the sample survey I helped devise and pre-test.

The sample survey's use of structured, fixed-choice questions without ethnographic research could not address the daily life obstacles and emotional conditions faced by my consultants. For example, what I perceived as changing self-conceptions as they expressed themselves about pending death, details about their irritations with their children, especially over financial matters, and sometimes on-going debates about a possible move to an institutional, assisted-care facility while facing the consequences of deteriorating bodily and mental faculties. The survey seemed like a constrained way to understand how retired persons with limited resources coped with daily life routines such as shopping, paying bills, and finding transportation for medical appointments that often became serious challenges.

Social activities at the Senior Citizen's club provided an opportunity to observe:

- sometimes paradoxical social exchanges at lunch where the interaction could be confusing, rather than smooth, connected discourse involving coherent topics. Instead, the discourse could be terminated without clear speech acts despite the initial appearance of 'normal' speech events between two or more persons;
- a fairly diverse set of members, some of whom had worked at a large factory, others had moved to the local small town from nearby rural areas, and included some former students from the nearby large university;
- social activities such as dancing and the existence, occasionally, of sometimes awkward appearing physical movements that could be interpreted as sexually motivated;
- interaction between persons of strikingly different economic, occupational and educational backgrounds that seldom, if ever, extended to evenings and weekends, including a few former doctoral students of

> the nearby university, and local, low and middle income persons who had never left the community.

After months of sometimes daily and frequent weekly contacts, I began to be invited to informal dinner parties at my consultants' homes. These informal occasions helped me recognize the relevance of ecological validity as a serious issue in social science research.

The brief fragments of descriptive material presented above were originally based on field notes taken while observing the social interaction of those studied, and by leaving the premises briefly to write down what I could remember while sitting in my automobile parked nearby. I did not have access to a tape-recorder and none were being used by other fellow graduate students in my joint anthropology-sociology department.

I am not, however, equating ecological validity with field research and the interpretation of ethnographically derived descriptive fragments. Instead, I call attention to my perceived disparity between the sample survey questions I helped devise based on then current theory and research methods, and my field work. The abstract nature of the theories used has not changed, nor have the methods (with the exception of the use of simulation studies).

The survey questions attempted to operationalize macro-level theories of social structure and were useful for formulating policies. What I felt was missing were the kinds of everyday practices associated with the problems retired persons encounter in their day-to-day activities. Constructing and asking questions of a diverse population while using the same language structure is a daunting challenge because it is difficult to assess the respondents' understanding or ignorance of the content of each question, despite the fact that they readily fill in fixed-choice questions about their ability to cope (in the present case) with life after retirement. For example, respondents can check that they are 'often' or 'sometimes' or 'never' lonely, but observing their daily life routines can provide a level of validity about the day-to-day circumstances that fill their daily lives. It is difficult to attain a clear sense of their daily lives by retrospective, abstract questions within a limited time span. The ways in which different respondents' memories seek to reconstruct their daily life activities and emotional experiences are topics that survey research can address but are limited by the method.

The reliability of questionnaires, therefore, can be deemed high with respect to the consistency of the responses to questions asked, but their validity is likely to be weak vis-á-vis understanding their daily life activities over weeks or months.

The ethnographic materials were based on my observation of repeated activities during a year of participation and observation of subjects at the Senior Citizens Center, interacting with them in their homes, and while directing a men's group on outings on Saturdays. As I noted above, many of those observed spontaneously reported negative feelings about their children (but not their grandchildren), their daily round of activities, dealing with illness, and meeting economic challenges at gatherings of their friends and at the Senior Citizens Club.

Formulating questions about the aforementioned issues in a survey is not difficult, but assessing their validity using the same instrument is problematic. Combining survey-type studies with ethnographic participation and observation remains illusive because many funding sources are not interested in systematically pursuing methodological problems unless they are convinced that the substantive results will be useful, for example, choosing a small, random subsample of respondents and re-interviewing them with open-ended versions of the original questionnaire and then accompany them on their daily round of activities for a week. The latter strategy does not qualify as 'ethnographic', but can add considerable validity to the original survey results.

In subsequent studies, I began to introduce audio and video recordings of some of the daily life settings I could observe in order to clarify activities, discourse, and socially organized interaction that are assumed to be inherent in, but not examined explicitly by the assembly and interpretation of aggregate data.

6. Inside juvenile justice

Although the ecological validity notion remained somewhat illusive and implicit rather than explicit in 1960, I pursued the notion by a four-year study of the administration of juvenile justice in two southern California cities of 100,000 persons in each (Cicourel 1968). The two cities appeared similar, but after the first year of field research, they seemed to possess rather contrasting city governments and police departments. One (City A) appeared to have 'normal' corruption while the other (City B) appeared to be especially corrupt. In City A, I was able to compare unofficial and official statistics compiled by the juvenile division and found that only the official records were sent to the Chief Statistician in Sacramento, California. During a subsequent interview with the Chief Statistician, he noted he was unaware of the 'double-bookkeeping' system in City A.

After spending many months accompanying plain-clothed detectives during the day, nights, and on weekends in City A, I realized the informal

records were a compressed representation of the juvenile officers' interaction with actual clients. I decided that a more extended research project would be needed and received encouragement from colleagues in a group in the Law and Society Center at the University of California, Berkeley law school (Erving Goffman, Ruth Kornhauser, Edwin Lemert, David Matza, Irving Piliavin, Sheldon Messinger, Jerome Scolnick, Philip Selznick, and Carl Werthman).

In City A, therefore, the same routine daily activities of the police generated the unofficial records kept by the juvenile division and the official police reports sent to Sacramento. What remained unclear was the process by which juveniles could accumulate 3–5 notations in the unofficial records before having a petition filed with the Juvenile Court. Filing a petition meant that the juvenile's activities became part of the official police statistics sent to Sacramento.

As I accompanied police officers (not all of whom were with the Juvenile Division), I concluded that the perception, apprehension and conviction of juveniles who allegedly broke the law included selective judgments often based on social class and ethnic biases. These biases appeared to be especially expressed in City B. Although no informal record system existed, each officer remembered their frequent encounters with specific juveniles.

The invariably terse descriptive remarks in actual police reports of the officers contrasted with the ways in which interrogations of juveniles were conducted during actual encounters. I also examined how juveniles were interrogated and judged at different stages of the juvenile justice system in the two cities, beginning, usually, with the police and sometimes ending in a court setting.

An appointment as a probation officer without pay in both counties enabled me to access large samples of official and unofficial records and observe the daily practices of police and probation officers, as well as observations at each city jail and juvenile detention facilities. I was also able to observe state correction facilities for juveniles.

The formal and informal records assembled by the police and probation departments, my experiences accompanying police and probation officers while they pursued their daily routines made it possible to explore aspects of official statistics ethnographically. If the primary source of data was the selectively aggregated information submitted by the police and probation departments to statewide agencies and subsequently compiled by the FBI in the U.S. Department of Justice, it would not be possible to observe the day-to-day activities by which official statistics were produced.

The field research enabled me to show that contrary to the official statistics on rates of delinquency by standard measures of social class and

ethnicity, there were approximately equal rates of infractions within middle/upper middle class families as there were within lower middle and low income families.

In City B, officers would often call the middle income juvenile's home before taking serious action such as taking the suspect to the city jail or to the juvenile detention facility. Or, in City A, only giving a warning to well-groomed adolescents creating a disturbance at a party in a neighborhood in which they resided.

In City B, a common practice appeared to be to arrest minority adolescents creating a similar disturbance in their own neighborhood, and stopping a minority adolescent for questioning if found in a neighborhood which appeared more 'upscale'. In City B, I observed detectives calling parents whose military officer father was stationed at a nearby air force installation after the children were arrested by the police. The detectives often asked such parents to come down to the police station to pick up their child. Detention and filing a petition for a court appearance was thus sometimes avoided.

In City A, parents from middle and upper middle income families often suggested that the juvenile who committed an infraction equal to an adult felony could be sent out of the county to a relative or to a private school. Officers often agreed, and a court petition was not filed. For the officers in City A, not having the juvenile remain in the city was economically advantageous and saved them considerable paperwork. Such arrangements are difficult to discern with sample surveys. Obtaining such information by direct observation required considerable, labor-intensive field research and informal on-the-job, open-ended interviews and follow-up probes.

Learning about the dual system of record-keeping in the less corrupt city (A), and obtaining permission to photocopy them, was only possible after more than a year of field research and gaining the confidence of officers. The dual system amounts to police-controlled 'probation' that somewhat paralleled the official Probation Department activities supervised by the county and the juvenile court.

I also was able to obtain large samples (500 from each city) of official statistics and, for example, tabulate the number of offenders and offenses, the number of repeated offenses, the total number of offenders in each city for particular years, among other statistics. A more labor-intensive task required hand copying of actual police reports in each city. I was able, therefore, to compare aggregated statistics and actual reports with field notes of my observations and eliciting information as officers pursued their duties.

I was not allowed to tape-record routine police/probation/officer-client interaction but I was able to obtain a few fragments which have been

presented in my book, as well as recording a long interview of a juvenile by a probation officer.

The juvenile justice study, therefore, strongly suggested that the notion of ecological validity in the social sciences could only become viable by linking it to the study of daily life social interaction in different formal and informal organizational and family or other settings. I concluded that whenever possible, discourse materials should always be part of the study of larger socially organized contexts. Speech acts, therefore, always reflect more complex speech events and the socially organized settings within which they emerge.

Within bureaucratically organized public and private institutions, information is aggregated by often unknown personnel using undocumented coding strategies. I was sometimes able to observe coding strategies by examining police and probation reports within the two cities and counties when such reports were being written, and being instructed by personnel on how these tasks should be accomplished. The major obstacle to this combined macro and micro research is obtaining access, always the key to attaining some approximation of ecological validity.

The issue of access was a continuous problem because some of the police officers in both cities were either suspicious or occasionally (but not always) hostile throughout the study. In the most hostile (and corrupt) city (B), the officer in charge of the juvenile division became chief of police during my last year of research and allowed me to listen to (but not record) his incoming telephone calls. In the city with 'normal' corruption, there was less hostility and, I think, fairly good relations. An anecdote may give the reader a sense of the validity of my findings and an object lesson to field researchers who think they have been 'accepted' by the 'natives.' I gave a copy of my book on the two cities to the group of detectives that pursued juvenile cases in the less corrupt city (A), and then had lunch with them at one of their homes. I was told that my findings were accurate but that if had they realized that I would be faithful about not revealing names and details about individual officers' activities, they would have given me more information about how the department functioned.

The last paragraph is, in my opinion, an example of both the value of field research and its limitations. Daily life participation in a group's activities does not guarantee 'acceptance' and 'trust' by the 'natives'. Continuous participation and/or observation, however, makes it difficult for subjects to mask their daily routines, but can also mean that they do not necessarily say what they mean, nor mean what they say when a visible outsider is listening and/or tape-recording their discourse.

My judgment of differential corruption in the two cities mentioned in this section is based on several types of activity. For example, in the case

of the 'corrupt' city, I learned that the Chief of Detectives resigned from the force because he was suspected of using money to pay police informants to settle his gambling debts in Las Vegas. One juvenile office spent considerable time running a small business at the annual County Fair from his police desk during a given year. Another officer (who became Chief of Police) used his police office to see clients during the income tax season. The county sheriff was known to have been involved in several serious illegal activities. The mayor was charged with allowing city workers to pave the parking lots of a few physicians who donated money to his election campaign.

In the 'less corrupt' city, I sometimes accompanied a detective who would regularly take the proceeds of his wife's small business to the bank while officially working.

The juvenile justice study, therefore, convinced me that aspects of ecological validity could be realized if hypotheses generated by compiling official rates of delinquency and linked to demographic information on each juvenile were compared to systematic observation of the daily life activities of personnel within the juvenile justice system. Juvenile justice could then be seen as emergent and patterned social processes rather than abstract social categories.

7. Ethnography and language use in classrooms

The third project addressed classroom and testing performances in the assessment of educational achievement in two elementary schools in Goleta and Santa Barbara, California (Cicourel et al. 1974; Mehan 1979). Considerable time and effort was required to obtain permission to study two first-grade classrooms in two schools. Before the second year of research could begin with the able assistance of several graduate students (Kenneth Jennings, Sybillyn Jennings, Kenneth Leiter, Robert MacKay, Hugh Mehan, and David Roth), I spent one academic year observing in a kindergarten and first-grade classroom in the Goleta school while trying to convince the two principles and two teachers (one from each school) to allow us to videotape their first-grade classrooms for one academic year.

I gave a lecture at each school to their faculties in the hope of gaining entrance to a classroom, but no one in either school volunteered. In the Goleta school, the kindergarten teacher (whose class I had observed) came to my rescue by 'lobbying' successfully for me with one of the first-grade teachers. In the Santa Barbara school, I finally obtained permission from a teacher whose husband was a colleague at UCSB. The teachers were, understandably, worried about having a camera present in the

classroom every day despite knowing that it would not be 'on' at all times.

A general goal was to understand the role of language use in classroom lessons and psychological testing, and how these activities were used to assess the educational performance and progress of European-American and non-European ethnic groups.

In addition to video recordings of classroom lessons and systematically scheduled observations within the classroom, we also videotaped official testing performances by school psychologists, as well as additional testing by the research team in the schools and in some of the children's homes. We also tested children on standard language acquisition issues such as their ability to understand the use of locatives when given oral sentences they were asked to act out.

The general idea was to compare children's performance on classroom lessons with testing results at school and at home. We concluded that minority children could perform better at home than at school (and in one case at home, even while watching a television program in another room). The standardized tests were divided into two parts; one part administered at school and the other at home. The Latino children whose first language was Spanish were tested in their native tongue at home.

The videotapes of classroom lessons and psychometric testing revealed the necessarily selective nature of teaching and testing; teachers are not always able to perceive details about children's behavior during a lesson, nor the kinds of distractions that children received from other children and/or initiated themselves. The teacher's assessment of a child's performance in the classroom, therefore, appeared to be linked to their memory of prior performances and an expectation of no change. If children were distracted or (we speculated) bored, as was more often the case for children identified as 'low achievers', they often did not follow the teacher's instructions. Labor-intensive observations, note-taking, and the videotapes allowed us to make such assessments and we believe contributed to the ecological validity of the study. Without a team of motivated researchers, such data would be difficult if not impossible to obtain and assess.

In the case of psychological testing, and based on difficult-to-document on-line observation and viewings of the videotapes, we discerned a tendency on the part of the testers to be more lax with children they knew were high achievers when waiting for an answer to a test question.

In our own testing, we found that allowing minority children to go beyond the cut-off point of three incorrect responses of the standardized Peabody Picture Vocabulary Test produced a significant number of correct answers in subsequent parts of the test. We also would ask the children to tell us about those pictures and words they could not identify.

For example, the word 'thoroughbred' was to be linked to a picture of a horse. When asked to identify the picture, he children could readily state the picture was that of a 'horse' and could do the same for a picture that was to be linked to the word 'shears' by immediately using the term 'scissors'. We speculated, therefore, that the teachers and tester's assessments were probably biased by tacit assumptions that minority children were less competent and the fact that they had not performed well on standardized tasks. For empirical details, see Cicourel et al. (1974) and Mehan (1979).

The ecological validity issue was approximated fairly well, but problems remained. For example, we had no way of assessing the validity of our sample of only two classrooms in two school districts despite our ability to videotape systematically across different days and times of day. The difficulty of obtaining teachers who would allow us to be intrusive was a serious problem vis-à-vis our ability to generalize our findings, as is the case in all field research. Another drawback was the difficulty of learning how parents and peer groups influence the children's motivation to learn, and the kinds of support that exist or are absent in the home.

Aggregate data (grades and test scores) on school performance does not include the process of student participation and teacher assessment on-line in the classroom and in testing settings; studying these processes required observing individual and group performances in daily life educational settings. Sample surveys and demographic data alone can be suggestive, but cannot tell us about the role of cognitive/emotional mechanisms and the social interaction and interpersonal skills that enable individuals and groups to confront the contingent processes that produce actual outcomes and how they affect educational career choices, obstacles, and subsequent occupational careers and their distribution across a population.

8. Conclusion

Perhaps the most important claim by W. S. Robinson was his remark that in social science research, reliability increases with sample size, but validity decreases. Combining the two is a serious conceptual and methodological issue that social scientists tend to avoid. The emphasis on reliability and creating generalizations are clearly essential research goals, but their achievement at the expense or neglect of ecological validity is usually justified by another important goal: making predictions. Virtually all predictions in the social sciences are demonstrated after the fact, except for experiments whose ecological validity is not explored.

One referee asked that I bring up to date my reliance on research on ecological validity conducted many years ago. A more recent paper (Ci-

courel 1996) addresses the ecological validity issue and includes different but relevant data. However, to my knowledge, the issue of ecological validity began within experimental psychology and though viewed as important, is seldom addressed systematically by academic psychologists. The primary focus remains on laboratory research, and, increasingly, devoted to cognitive neuroscience explanations and the use of electrophysiological measures like Evoked Potentials and functional Magnetic Resonance Imaging.

For some 25 years after the research reported above, I was able to approximate ecological validity issues by studying physician-patient communication and diagnostic reasoning in a school of medicine. I held regular academic appointments in cognitive science, our medical school, and sociology. As a faculty member in the School of Medicine (primarily in pediatrics), I taught undergraduate medical students and residents, and observed their interaction with patients. In pediatrics, I helped train all first-year residents by videotaping their interviews of mothers and their children and the way they conducted physical examinations. I reviewed the tapes with each resident during a four-week rotation.

My frequent professional and social contacts with the medical faculty enabled me to study activities in different clinics and hospital laboratories as an 'insider'. I published many papers on my medical research and was fortunate in having considerable access to many different settings. I was able to tape-record while observing social interaction and at times interviewing medical personnel while they were engaged in specific tasks (Cicourel 1982, 2003)

The medical school research could be viewed as approximating the years that biological behavioral ecologists spend studying non-human animals in the wild (Krebs and Davies 1997). We seldom study human groups for such long periods of time (but see Whiting 1966, and especially Hanks 1990). Research funds for longitudinal social science research are seldom awarded. Such studies require labor-intense effort and are often termed 'high risk' because of delays in publishing steadily. The trade-off between research for a 'high risk' book and a 'low risk' journal article and/or anthology publication varies with academic departments, but cannot be ignored.

My field research cannot be equated to the traditional (e.g., pre-World War II) research in small villages in which the observer is able to perceive and/or participate in continuous day-to-day activities. Yet within communities and large bureaucracies, it is possible to go beyond a reliance on sample surveys and/or structured interviews by engaging in labor-intensive field research in order to approximate ecological validity. For example, in my study of juvenile justice, I was able to combine the

study of official and unofficial statistics and supplement my findings with detailed observations of the same organizations and personnel over a period of four years.

I noted above that I am pessimistic about funding agencies that would devote their resources to exploring the trade-off of different methods: for example, combining survey and/or demographic methods with labor-intensive field methods by asking teams of researchers to join forces and both conduct a survey of a fairly large sample of respondents first before seeking to observe and interact with a random, smaller subsample for a period of months. This strategy would be especially helpful in the study of routine healthcare delivery and the use of epidemiological/public health research across diverse populations.

Finally, I (like many others) found it difficult to spend four years studying juvenile justice systems in two counties and their largest cities, but also rewarding with respect to obtaining information from different consultants and observing the same personnel under different work circumstances.

The 25 years I spent in a school of medicine as a professor who taught both medical students and first-year residents enabled me to function as an 'insider' on many occasions when observing and recording physician-patient interaction. Many medical school colleagues were helpful by asking other colleagues I did not know to allow me to engage in research in their clinic. Having friends among the medical school faculty enabled me to obtain their assessments of my research findings. Over a period of several years, I have collaborated with a French colleague, Saadi Lahlou, in a human-computer cognitive design laboratory in Paris. I was the 'outsider' and Lahlou, as director of the Lab, the obvious 'insider'. We are currently drafting a paper on this research. The strengths and weaknesses of the 'insider' and 'outsider' often became painfully obvious to us.

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behavior* 49: 227–267.

Boas, F. (1938). *General Anthropology*. New York: Heath.

Bronfenbrenner, U. (1977). Toward am experimental ecology of human development. *American Psychologist* 32 (7): 513–531.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*, 2nd ed. Berkeley: University of California Press.

Cicourel, A. V. (1968). *The Social Organization of Juvenile Justice*. New York: Wiley.

—(1974). *Theory and Method in a Study of Argentine Fertility*. New York: Wiley Interscience.

—(1976). Representation and meaning and the study of deaf children. In *Life Sentences: Aspects of the Social Role of Language*, R. Harré (ed.), 146–159. New York: Wiley.
—(1977). Sociolinguistic aspects of gestural sign language. In *Current Trends in Studies of the Sign Language of the Deaf*, I. Schlesinger and L. Namir (eds.), 271–313. New York: Academic Press.
—(1982). Language and belief in a medical setting. In *Contemporary Perceptions of Language: Interdisciplinary Dimensions*, H. Byrnes (ed.), 1–41. Georgetown, VA: Georgetown University Press.
—(1992). The interpenetration of communicative contexts: Examples from medical encounters. In *Rethinking Context: Language as an Interactive Phenomenon*, A. Duranti and C. Goodwin (eds.), 291–310. Cambridge: Cambridge University Press.
—(1996). Ecological validity and 'white room effects': The interaction of cognitive and cultural models in the pragmatic analysis of elicited narratives from children. *Pragmatics and Cognition* 4 (2): 221–264.
—(2003). On contextualizing applied linguistic research in the workplace. *Journal of Applied Linguistics* (Special issue edited by Chris Candlin and Srikant Sarangi): 360–373.
—(2006). Cognitive/affective processes, social interaction, and social structure as representational re-descriptions: Their contrastive bandwidths and spatio-temporal foci. *Mind and Society* 5: 39–70.
Cicourel, A. V., Jennings, K., Jennings, S., Leiter, K., MacKay, R., Mehan, H., and Roth, D. (1974). *Language Use and Classroom Performance*. New York: Academic Press.
Cicourel, A. V. and Kitsuse, J. I. (1963). *The Educational Decision-Makers*. New York: Bobbs-Merrill.
Cole, M., Hood, L., and McDermott, R. P. (1982). Ecological niche picking. In *Remembering in Natural Contexts*, U. Neisser (ed.), 366–373. New York: W. H. Freeman.
Hanks, W. F. (1990). *Referential Practice: Language and Lived Space among the Maya*. Chicago: University of Chicago Press.
Krebs, J. R. and Davies, N. B. (1997). *An Introduction to Behavioural Ecology*. Oxford: Blackwell.
Loftus, E. F. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.
Mandelbaum, D. G. (ed.) (1949). *Selected Writings of Edward Sapir in Language, Culture, and Personality*. Berkeley and Los Angeles: University of California Press.
Mehan, H. (1979). *Learning Lessons: The Social Organization of Classroom Instruction*. Cambridge, MA: Harvard University Press.
Murdock, G. P. (1967). *Ethnographic Atlas*. Pittsburgh, PA: University of Pittsburgh Press.
—(1975). *Outline of World Cultures*. New Haven, CT: Human Relations Area Files.
Neisser, U. (ed.) (1982). *Remembering in Natural Contexts*. New York: W. H. Freeman.
Robinson, W. (1950). Ecological correlatioins and the behavior of individuals. *American Sociological Review* 15: 351–357.
Schegloff, E. A. (1987). Between micro and macro. In *The Micro-Macro Link*, J. Alexander (ed.), 207–234. Los Angeles: University of California Press.
Whiting, B. B. (ed.) (1966). *Six Cultures*. New York: Wiley.

Aaron V. Cicourel is research professor of cognitive science at the University of California, San Diego. He has been a professor at UCSD since 1970. Most of his research over the past thirty years has been in medical settings as professor of pediatrics and simultaneously as professor of cognitive science and sociology. Address for correspondence: University of California, San Diego, Department of Cognitive Science, 9500 Gilman Drive—MC 0515, La Jolla, CA 92093-0515, USA ⟨Cicourel@cogsci.ucsd.edu⟩.