INTERVIEWER AS INSTRUMENT

Accounting for Human Factors in Evaluation Research

JOEL H. BROWN
Center for Educational Research and Development and the University of Oklahoma

This methodological study examines an original data collection model designed to incorporate human factors and enhance data richness in qualitative and evaluation research. Evidence supporting this model is drawn from in-depth youth and adult interviews in one of the largest policy/program evaluations undertaken in the United States, the Drug, Alcohol, and Tobacco Education evaluation (77 districts, 118 schools). When applying the explicit observation technique (EOT)—the strategic and nonjudgmental disclosure of nonverbal human factor cues by the interviewer to the respondent during interview—data revealed the observation disclosure pattern. Here, respondents linked perceptions with policy or program implementation or effectiveness evidence. Although more research is needed, it is concluded that the EOT yields richer data when compared with traditional semistructured interviews and, thus, holds promise to enhance qualitative and evaluation research methods. Validity and reliability as well as qualitative and evaluation research considerations are discussed.

Keywords: *evaluation methods; human factors research; interaction analysis; discourse analysis*

In qualitative evaluation research, there is an inherent interest in enhancing data richness to enhance study quality. With the opportunity to develop rich data often comes an interest in accounting for human factors among study participants. The goal of the following research is to explore an interview model designed to include and account for many such human factor dimensions of the qualitative interview. Specifically, this methodological article shows how in a large-scale evaluation, researchers used nonverbal

---

AUTHOR'S NOTE: *No part of this article may be reproduced in any form without the author's express written consent. Portions of this research were supported by the California State Department of Education Contract #3279. The views expressed herein are those of the author and do not necessarily represent those of the California State Department of Education. Special thanks to David Brown and Sylvia Thyssen for editorial assistance.*

EVALUATION REVIEW, Vol. 30 No. 2, April 2006 188-208
DOI: 10.1177/0193841X05277222

cues as a basis for interview probes. Theoretically, it is asserted that addressing such cues during interview allows researchers to account for key human factors and connect such data with evaluation phenomena—thus enhancing study quality.

Evidence regarding this approach comes from the California Drug, Alcohol, and Tobacco Education (DATE) evaluation, one of the largest mixed-method educational evaluations conducted in the United States (77 school districts, 118 schools, more than 5,000 young people). It has yielded numerous publications, as well as receiving recognition from the National Academy of Sciences (Manski, Pepper, and Petrie 2001; Brown and D'Emidio-Caston 1995; Brown, D'Emidio-Caston, and Pollard 1997; D'Emidio-Caston and Brown 1998). With the following research, the yield of this remarkable data set continues.

The focus of this article is not on the study results per se but rather on the use of methods leading to the discovery of the previously published evaluation outcomes based on such procedures. As such, several terms are defined and referred to throughout the article:

1. The technique eliciting "richer" data is referred to as the explicit observation technique (EOT), defined as the interviewer explicitly and systematically noting with respondents' observations of nonverbal cues as a strategic adjunctive probe to scheduled interview questions during a qualitative interview.
2. *Nonverbal* is defined by the dictionary as "being other than verbal; not involving words" (American Heritage Dictionaries 2000).
3. Richer data are seen as evidence helping to account for interview human factors and assess policy or program implementation or effectiveness (Schutz, Walsh, and Lehnert 1967; Spradley 1979; Glaser and Strauss 1967; Mischler 1986).
4. Evaluation phenomena are defined as respondent data specifically describing incidents and/or perceptions regarding policy or program implementation or effectiveness.

Based on these terms, the main assertion is that using the EOT will yield a richer data set than might be found in traditional qualitative research and/or evaluations.

The article develops in the following way. Following this introduction, a brief research background is explored. This exploration is succeeded by the DATE and EOT research methods and presentation. Exemplars representative of the qualitative evaluation findings are presented. The discussion focuses on summarizing findings, reliability/validity, and evaluation conclusions.

RESEARCH FOUNDATION: INCORPORATING HUMAN FACTORS IN QUALITATIVE EVALUATION

In his classic book *Research Interviewing*, Mischler (1986) microanalyzed the psychological mechanism(s) of human factors during the qualitative interview. A basic assertion is that accounting for such human factors allows for a potentially richer data yield than do those using traditional structured or even semistructured interview schedules. During the interview, he suggested that such an enriched interaction occurs through the

> mutual reformulation and specification of questions by which they take on particular and context-bound shades of meaning between interviewer and respondent. . . . [This] helps develop an interview whereby the context of the question is given specifics by the respondent and the answer is accepted by the interviewer. (P. 53)

Traditionally, these human factors have been accounted for with methods including "thick description" of the interaction (Geertz 1973) or detailed note taking during or after the interview. Even in their most advanced forms, when using such methods, determining the meaning of human factors such as nonverbal cues and their connection(s) to the interview situation, policy, or program may still be insufficiently understood, let alone incorporated into the qualitative interview.

What is needed is a method of translating Mischler's theory into an effective practice of observing and accounting for such human factors during the interview. Horowitz and Brown (1996, 115) asserted that a means to accomplish such accountability in the service of richer evaluation data is by "making explicit those elements implicit" during an interview; specifically, they suggested that systematically noting such factors to the respondent during the interview "allows a new understanding to emerge" and "advances the methods and interpretation of evaluation research findings."

Judith Brown saw the key means of "making explicit those [human factors] elements implicit" as interviewer training. To achieve this, she and others (Berg and Smith 1988) found a need for the "conscious self-involvement" of the interviewer with the respondent. Brown further noted, "Although the recent literature in research methods—in qualitative methodologies—is vast and the need for training for conscious self-involvement is frequently mentioned, any concrete suggestions about to accomplish this are difficult to find" (Brown 1996, 1).

The following research, then, extends Mischler's theory regarding the importance of incorporating "context bound shades of meaning" into the

qualitative interview. This is developed by doing as Judith Brown and Horowitz and Brown (1996) suggested, training the interviewer to be more self-conscious during the interview by learning how to systematically consider human factor cues in real time. The purpose of this research, then, is to determine if methods designed to do this produce a richer data set than might be found using traditional methods.

METHOD

OVERVIEW

The evidence presented in this methodological study comes from the DATE evaluation, originally conducted between 1991 and 1994 for the California State Department of Education. The goal of the evaluation was to determine program implementation and effectiveness of drug education services delivered across California. The cost of the overall evaluation was approximately $4 million. Participation in the evaluation was mandated by the state of California.

The DATE evaluation included extensive qualitative and quantitative data collection and analysis with adult and youth school system participants. The quantitative portion of the evaluation was conducted in 118 schools in 77 school districts. It included extensive cost and implementation analyses, as well as surveys with approximately 5,000 randomly selected students in Grades 7 through 12.

The goal of the qualitative portion of the evaluation—drawn from a subset of 50 of those districts—was to determine program implementation and effectiveness while describing and understanding the social world of program delivery participants and program recipients (Schutz, Walsh, and Lehnert 1967). Although presented in full in earlier publications, for each of the 2 years of data collection and analysis, an overview of the qualitative methods related to this methodological study is presented.

The interview technique discussed in this article was discovered by our team during pilot qualitative data collection. One researcher consistently yielded richer data (as previously defined) from those interviews. While following the semistructured interview format, during probes, he occasionally reflected the respondents' nonverbal cues by making selected ones, such as facial expressions, explicit. Given that several team members were also clinical professionals, all interviewers were subsequently trained to determine if,

while fulfilling all human subjects requirements, they might also elicit richer data. Following the analysis and publication of the DATE evaluation findings themselves, subsequent analyses were conducted to examine the yield of EOT versus non-EOT interview data. This methodological article, then, focuses on the serendipitously discovered and intentionally applied aspect of qualitative evaluation.

YEAR 1 METHODS

To ascertain program implementation type and level, the 1st year of qualitative evaluation focuses on those developing and delivering DATE drug education services.

PARTICIPANTS

District and school interviews were conducted in the field with seven to nine key personnel, from 108 schools in 50 school districts (LeCompte and Preissle 1993; Zelditch 1962; Spradley 1979; Gilchrist 1992). Included among the seven to nine key personnel targeted for interviews in each district were the district DATE coordinator, one DATE coordinator supervisor, one DATE coordinator staff member, the district financial coordinator, the superintendent or assistant superintendent, and the DATE site coordinator. A teacher at each of 2 schools visited and a community member involved with the DATE project were selected by the district DATE coordinator, and they were scheduled for interviews as well.

INSTRUMENT

A broad interview schedule was used as a semistructured instrument to determine the social influence processes used in drug education programs and to ascertain program implementation and effectiveness. The instrument is available in Brown and D'Emidio-Caston (1995).

DATA COLLECTION

The left-hand portion of Figure 1 provides an overview of the Year 1 data collection and how a representative data sample was derived from the

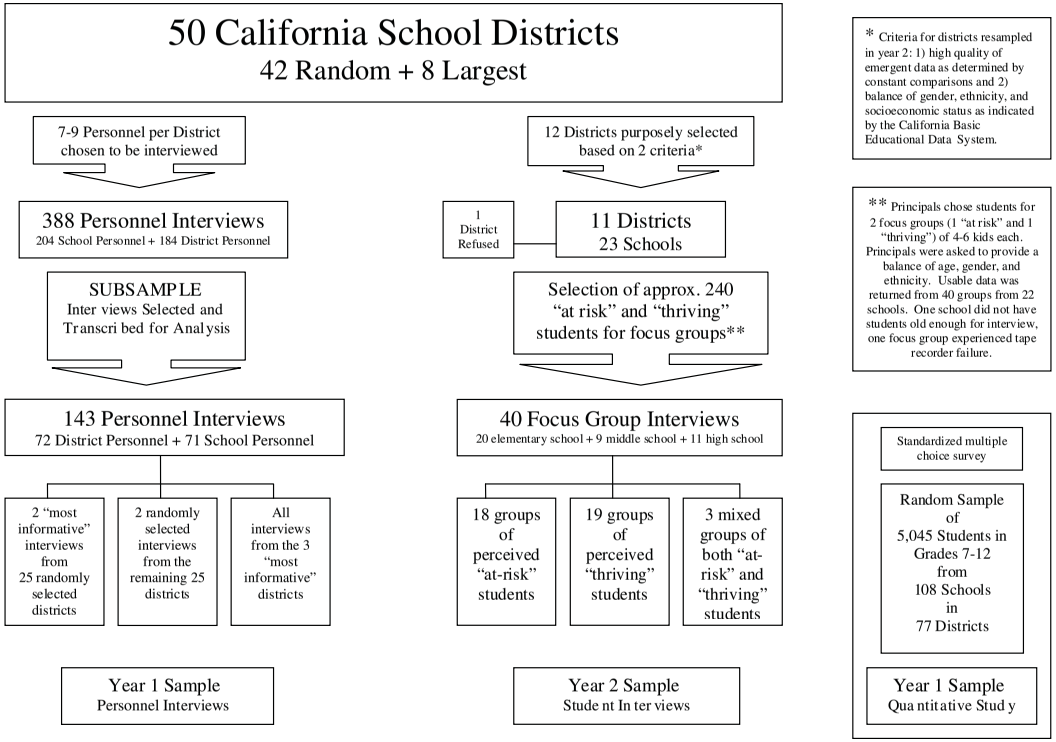

Figure 1: Overview of California Drug, Alcohol, and Tobacco Education Evaluation Sampling

collected data. The qualitative portion of the evaluation included in-depth interview data collected from 50 school districts (the 8 largest districts and 42 other randomly selected districts). Three schools were randomly selected from each of the 8 largest districts, and 2 schools were randomly selected from each of the 42 smaller districts, resulting in 108 schools. If a district had only 2 schools, both were included.

The interviews occurred in a private, secluded area without the presence of any other individuals and lasted between 30 and 60 minutes. Interviews took place over a 2-day period and were audiotaped after the participants gave informed consent.

Of the collected 388 interviews (7-9 personnel from 50 districts), 143 were selected for analysis. From each of the 25 randomly selected districts, the 2 most informative interviews, as determined by the interviewers, were purposely selected (Marshall and Rossman 1989). From each of the remaining 25 districts, 2 interviews were randomly selected for analysis. Finally, all interviews from the 3 "most informative" districts (selection based on previous interview data) were selected for analysis. This process resulted in the transcription and analysis of 72 interviews from school district personnel and 71 from school site personnel, providing a sampling balance among organizational sites and levels.

YEAR 2 METHODS

To ascertain program impact, the 2nd year of qualitative evaluation shifted from school and program personnel to program recipients, the students receiving DATE programs.

PARTICIPANTS

Forty focus group interviews with nearly 250 young people were intentionally drawn from 12 of the previous year's school districts. Focus groups were composed of approximately 4 to 6 students, Grades 5 through 12. Under theory regarding perceived policy requirements regarding at-risk youth, as selected by the principal of each selected school, focus groups were developed based on specific "at-risk" versus "thriving" evaluation criteria as established in Brown and D'Emidio-Caston (1995). The principals were asked to provide a balance of gender, ethnicity, and grade level. Researchers were not allowed to formally record demographics because of strict anonymity limitations from the Department of Education.

INSTRUMENT

Two developmentally appropriate semistructured interview schedules were devised, one for Grades 5 and 6 and one for Grades 7 through 12. Schedules were devised to examine the social influence processes used in drug education and to determine student perceptions of program efficacy. Interviews lasted between 30 and 60 minutes. Actual questions can be found in Brown and D'Emidio-Caston (1995).

DATA COLLECTION

The right-hand portion of Figure 1 provides an overview of the Year 2 data collection and how a high-quality sample was derived from the collected data. Twelve school districts from the 1992 sample were drawn for resample based on two criteria: (a) a high quality of emergent data from each district as determined by constant data comparisons and (b) a balance of gender, ethnicity, and socioeconomic status as indicated by the California Basic Educational Data System (California State Department of Education 1992). Of the 12 selected districts, 1 declined to participate. Of the 11 sampled districts, 1 was among the state's 8 largest districts. Three schools were randomly selected from the large district, and 2 schools were randomly selected from each of the remaining districts (if a district had only 2 schools, both were included). In sum, 23 schools reflective of the state population of schools were selected for student interviews. Usable focus group data from the 40 focus groups as described above were returned from 22 of those schools.

The student focus group data included 18 perceived "thriving" groups, 19 perceived "at-risk" groups, and 3 "mixed" groups (perceived thriving combined with perceived at risk). The 3 mixed groups were from the largest school district and offered a means to compare data with at-risk and thriving groups. Twenty focus groups from 10 elementary schools, 9 groups from 6 middle schools, and 11 groups from 6 high schools formed the data corpus of 40 student groups. By combining our research goals with California's student population characteristics statistics, we sought to achieve a sample representative of California's school districts, schools, and students.

EOT TRAINING

Each of the 11 interviewers possessed extensive experience in qualitative interview and program evaluation. These skills were bolstered by approximately 24 hours of training specific to the adult and youth interview

schedules and the EOT. As adjunctive probes, the EOT training was designed to make explicit elements of the respondent's human factors. Interviewers engaged in a number of training activities overseen by the principal investigator, who in addition to program evaluation possessed extensive experience as a clinical social worker. He and his staff implemented training activities focusing on

1. consciously observing verbal as well as nonverbal human factor cues and, when appropriate, noting observations with respondents and
2. making primarily descriptive and nonevaluative statements during such probes.

Specifically, training activities focused on teaching interviewers to listen to what people said and how they said it. They learned how to observe and, as appropriate in their professional determination, conduct adjunctive probes by occasionally making explicit respondent nonverbal cues during the interview, such as facial expressions. For example, if the interviewer observed that the respondent changed his or her facial expression in response to what the interviewer determined was an important juncture during the interview, the interviewer could reflect this observation back to the respondent in a descriptive and nonevaluative way; for example, "you smiled when I asked this question." "Descriptive and nonevaluative" is seen as the interviewer noting only the observation of the behavior absent any interpretation of meaning with the respondent. Theoretically, such observations serve to invite the respondent to interpret the meaning of what was observed and noted in real time. Within the semistructured interview then, interviewers were trained to conduct adjunctive probes under this condition: when, in the professional discretion of the interviewer, an adjunctive probe would help account for the interviewer-respondent relationship or clarify the respondent's intended meaning regarding program implementation or effectiveness. For more specific information regarding training techniques, please see Brown, D'Emidio-Caston, and Benard (2000).

YEAR 1 AND YEAR 2 DATA ANALYSIS

For both Year 1 and Year 2 data, the research team consisted of seven members selected for their expertise and their varied backgrounds. The team included two quantitative educational evaluators, one public health substance use specialist, two substance use/abuse researchers, one educational/organizational researcher, and one substance use and educational/organizational researcher.

Identical analytical methods were used for both years, with one exception. In Year 2, additional quantitative content analysis was performed, in which student statements were logged and categorized (Berelson 1952).

The data analysis was based on the theoretical approach of symbolic interaction. From this perspective, cultural meaning is constructed through shared definitions of reality (Blumer 1969). Blumer (1969) hypothesized that to understand the meaning of any social phenomenon, it is critical to gain insight into the patterns of interaction that define the social world of the participants. By asking people to explain their perceptions of critical social processes, the grounded theoretical approach uncovers the meaning of shared interaction, the taken-for-granted reality of the culture under study (Glaser and Strauss 1967; Garfinkel 1969; Berger and Luckmann 1967; Strauss and Corbin 1990).

Data analysis was conducted at the district, school, and student levels using the constant comparative method. Using this method, researchers constantly compare respondent statements within and between interviews to determine similar and/or dissimilar statements of beliefs and behaviors of those involved in the development and implementation of the DATE program. Through rigorous categorizations of such statements, researchers gained an in-depth understanding of what DATE services were being implemented, why they were being implemented, and how they were being implemented. In addition, through constant comparisons, researchers tested the assertion presented at the outset of this article—that the EOT would elicit richer data than would non-EOT data as previously defined. This analytical method was designed to allow assertions to emerge and evolve as data were compared, ultimately resulting in a set of findings grounded in data.

Findings are supported by exemplars drawn from the data set. Each exemplar meets the criteria of inclusion determined by defining characteristics of each topic or category, as drawn from constant comparisons of transcribed interviews, interviewer's comment sheets, and field notes. Initially, each researcher analyzed data independently. Then, approximately every 2 weeks, the qualitative team met for 2 days to go over their own assertions and to compare them with other group members' findings. Through researcher interaction during these meetings, consensus was formed about the meaning of the social world as well as the similarities and differences between EOT versus non-EOT information represented in the data, thus arriving at this article's qualitative findings.

Overall, evidence was considered valid only after contradictory information, spurious relations, and rival explanations were considered and subsequently accounted for (Sanders 1994). Unless otherwise stated, exemplars presented represent the conclusions drawn from this examination of the evidence.

FINDINGS

During Year 1, the goal was to determine the scope and magnitude of program implementation. To ascertain program impact, the 2nd year of qualitative evaluation shifted the focus to examining the social world of young people under DATE. For a presentation of full empirical results, please see the previously referenced publications. Year 1 and Year 2 findings are presented in tandem.

In this first exemplar, special importance is ascribed to the respondent's facial expression. Using the EOT, the interviewer notes an observation to the respondent, and subsequently, the course of the interview interaction appears to evolve. Particularly germane statements appear in boldface:

> Interviewer (I): What do you feel are the differences between what the [school] district wants done, what the schools implement and what the students experience? **You're smiling**.
>
> Respondent (R): **I'm smiling because I think we have such a little concept of what the district wants done with the DATE money**. Nobody has come and said this is our vision for the district to do programmatically with drug, alcohol and tobacco monies. I've never seen that. When I came here last year that was, probably, my biggest shock, there were no programs, there was nothing going on this campus. If a kid had a drug or alcohol or any kind of substance abuse related problem, I had no place to refer them. There were no intervention programs going on, there were no prevention programs going on, there was no peer counseling. V [a person's name], I never saw last year. (interview no. 230, 8 [page number from transcript of interview])

During administration of an interview schedule item, the respondent significantly altered his or her facial expression as a key scheduled question was asked. To the interviewer in that moment, the smile was the most revealing aspect of the interaction. Perhaps by noting "you're smiling," the respondent became aware of information that she or he may not have been aware of. In concert with adjunctive probe criteria, this indicated to the interviewer that the interaction may be at a juncture in which the data richness may evolve depending on the interviewer's next statement.

The interviewer then feeds this observation back to the respondent in a descriptive rather than evaluative way, simply noting the smile. In so doing, he or she makes explicit an observation of a nonverbal cue. This appeared to yield an important data pattern.

Following the interviewer noting an observation, the respondent typically offered interpretive data linked with supporting program evaluation implementation or effectiveness evidence. In the above passage, the smile is noted, and a respondent interpretive statement, "I think we have such a little concept

of what the district wants done with the DATE money," is followed by several statements regarding program implementation and effectiveness. Researchers noted a linking statement between this disclosure and the program implementation: "Nobody has come and said this is our vision for the district to do programmatically with drug, alcohol and tobacco monies. I've never seen that." This statement closes with a direct statement about program implementation and potential for effectiveness: "There were no intervention programs going on, there were no prevention programs going on, there was no peer counseling." Through constant comparisons, our team coded this as contributing to pattern data. It provides evidence of the respondents' pattern of interaction around implementation of these services in his or her social world.

With respect to reducing potential researcher bias, rather than the meaning of the statement being primarily interpreted by the researchers, it is important to note that this evidence is largely interpreted by the respondent. When the EOT was used in the DATE evaluation, an emergent data configuration was located. The apparently resulting pattern is referred to as the "observation disclosure pattern," in which an interviewer's adjunctive probe elicits statements regarding intrapersonal perception linked with evaluation evidence regarding program implementation or effectiveness. Such evidence was consistently found to directly assist researchers in determining program implementation and effectiveness.

In addition to the presence of the observation disclosure pattern, when analyzing evidence, it is "of proven theoretical relevance" to consider patterns that are "notably absent when comparing incident after incident" (Strauss and Corbin 1990, 177). Based on this concept of "proven theoretical relevance" and systematic data presence and absence, in both the adult and youth data, researchers discovered a pattern revealing the absence of the observation disclosure pattern when the EOT was not used. For example, note the response to the identical question from the first passage in another interview on virtually the same organizational level:

> I: What would you say is the difference between what the District wants done, what the schools implement and what the students' experience?
>
> R: Well, I think the District and the school all want the same thing, they want to teach the substance abuse programs, they want to teach what these things do. I'm not sure what the students experience, that is almost impossible for me to say what they experience when they get to be teenagers. I certainly encourage all of the elementary drug education programs that are there. (no. 007, 7)

Our team found this information to constitute additional evidence for coding. For example, the respondent states, "They want to teach the substance abuse

programs" and that she is "not sure what the students experience." The passage is concluded with a general statement: "I certainly encourage all of the elementary drug education programs that are there." Some researchers such as Watzlawick, Weakland, and Fisch (1974) or Brown (1996) might have called the above nonsense data because the respondent apparently feels required to answer but may prefer not to, so she or he provides data that can be used for qualitative coding at a lower level than might be found under other circumstances. Nevertheless, in light of the absence of the kinds of rich evaluation responses—frequently the case when the EOT was not used across adult and youth data—this statement and other typical ones provide program implementation and effectiveness evidence to help saturate categories and themes. So, although providing useful categorical and thematic data for saturation and interpretation, such traditional qualitative interview methods were found to set the stage for richer data so often found in the EOT-related responses.

When using the EOT—this time noting body movement rather than facial expression—the observation disclosure pattern was also in evidence in the student data. In this passage, use of EOT appears to lead to important distinctions regarding the interpreted nature of zero-tolerance policy enforcement, defined as suspending or expelling a student from school for use, possession, or distribution of alcohol, tobacco, or drugs:

I: Is that the primary way they deal with it here? Suspensions and expulsions? **You're shaking your head and they can't hear you shaking your head!**

R: Yeah! That's how they deal with a lot of it. Related to drugs, alcohol and tobacco.

I: Any time they use?

R: Yeah, just about.

I: Or is it any time they use on campus?

R: **Well, I know a girl who didn't even go to this school. She went to W. and she came here and it was after school, she was picking up one of her friends. She was in the parking lot leaving school and was smoking a cigarette. They told her she couldn't come back to this school for so long! She wasn't allowed to come on campus! Even if she's 18 and allowed to smoke, they won't let her come back on to campus. What happens if your parent comes to school and he smokes? Your dad comes here and smokes on campus, they can't do that to him, so why could they do that to her?** (no. 506, 22-23)

Here, the EOT is used to descriptively note body movement during the interview: "shaking your head." This adjunctive probe leads to evaluation phenomena. In this case, the evidence includes a respondent description regarding a perception of how the school policies are enforced or implemented.

On one level, this yielded rich and important evaluation information, helping to distinguish the category of policy implementation or enforcement, where it was found that such policies "exclude those whom students themselves recognize as most in need of help and they drive those who already might be on the margins of the school system further out" (Brown and D'Emidio-Caston 1995, 482).

On another level, equally essential to the evaluation findings themselves is what emerges near the end of the respondent statement:

> Even if she's 18 and allowed to smoke, they won't let her come back on to campus. What happens if your parent comes to school and he smokes? Your dad comes here and smokes on campus, they can't do that to him, so why could they do that to her?

An important distinction emerges from constant comparisons in these data. This student draws a distinction between policy consistency: that of a young person of legal age smoking a cigarette and being prohibited from campus versus the same behavior from a nonenrolled adult being acceptable. As our earlier research demonstrated the importance of story in evaluation evidence (D'Emidio-Caston and Brown 1998), the veracity of the story may not be as important as what the disconcerting distinctions mean from an evaluation perspective—one of the most important findings from the DATE evaluation. In an educational pedagogy defined as disintegrative shaming, in which the educational objective is to "scare or shame youth into abstinence" (Brown 2001, 101), the

> qualitative evidence suggests that cognitive dissonance (Festinger 1957; California State Department of Education 1992) is linked with student descriptions of a state of tension. . . . Many students appear to resolve their cognitive dissonance by linking their perception of drug education with the new cognition that educators were lying to them about the information they provided or were not interested in helping those students they perceived as having a substance abuse problem. ["They lie to you so you won't do it!"; "I don't think the schools are for like helping it's just for getting the bad kids out."; #531, 21] Our survey results are also consistent with this contention. (Brown, D'Emidio-Caston, and Pollard 1997, 79)

From an evaluation perspective, the students provided clear and consistent evidence of DATE program (in)effectiveness. The cognitive dissonance associated with drug education programs such as Drug Abuse Resistance Education (D.A.R.E.) and Life Skills Training (LST) occurred through what the students viewed as inconsistency in what they were told about drug programs or policies versus what they observed or experienced. The dissonance was resolved in a reduction of educator/adult credibility, both in drug education and likely in general (Brown 2001).

From a methodological perspective, during the interview the use of the EOT as an adjunctive probe led to the development of a higher level finding. And here, typical of the observation disclosure pattern, interpretive data ("Yeah! That's how they deal with a lot of it. Related to drugs, alcohol and tobacco") are linked with supporting program implementation or effectiveness evidence ("Your dad comes here and smokes on campus, they can't do that to him, so why could they do that to her?").

Understanding the attitudes of program recipients appears to enhance the interpretive power when evaluating program effectiveness information. In this EOT interview, the student's response reveals negative attitudes toward program administrators and gives data on how DATE program recipients perceive the program's effect on themselves.

I: Do you like them teaching you this stuff or do you not like it? **You're shaking your head no?**

R: **I don't. Because they can say it all they want, but I'm gonna do it!** (no. 505, 35)

The use of the adjunctive probe makes explicit the student's body movement ("You're shaking your head no?"). The respondent then provides evidence regarding his or her attitudes toward program deliverers ("they") and affirms an assertion of program ineffectiveness by asserting his or her decision to use, despite the program. Again, we find the link of the adjunctive probe with evaluation phenomena, thus providing richer data than were typically found absent the probe. Other students in the focus group interview responded as well, making clear statements about program effectiveness.

R: They waste time! (*laughs*).

R: Yeah! 'Cause everyone does it anyways! (*laughs*.) Not everybody, but most people.

I: So, are you saying that no one should do anything? They should just leave you alone and not tell you anything?

R: They can do all they want, but they are just wasting their time! (no. 508, 2-4)

In another example, the use of the EOT precedes the observation disclosure pattern, leading from an interpretive statement to clear program evaluation data about "Red Ribbon Week." In this case, the student addresses the impact of the program on students who already smoke and/or drink.

I: What do you think of that? **You're smiling and laughing again?**

R: I don't know. What do I think of Red Ribbon Week? I think everyone gets a red ribbon from their class and wears it. **If the people that are using alcohol and wearing their red ribbon, just 'cause it's Red Ribbon Week they're not going to stop**

drinking! You're not going to stop smoking just because it's Red Ribbon Week! So, it's one week and I don't think it makes that big a difference at school. (no. 506, 26)

The use of the EOT is revealed as the reflection of the nonverbal reaction (smiling and laughing) without evaluating it. Following the use of the EOT, the observation disclosure pattern emerges as the student makes an interpretive statement ("I think everyone gets a red ribbon from their class and wears it"), which then leads to two very clear statements about program effectiveness. First, there is a statement specifically about the effect of the program on students who already use alcohol and tobacco: "Just 'cause it's Red Ribbon Week they're not going to stop drinking! You're not going to stop smoking just because it's Red Ribbon Week!" Next, the student makes a statement linking the program's structure to ineffectiveness: "So, it's one week and I don't think it makes that big a difference at school."

The absolute veracity of the students' statements is not as important as the impact that the students describe as linked with the program(s). These attitudes, thoughts, and feelings, emerging as rich information after the use of the EOT, yield pattern data for evaluating the implementation and effectiveness of these programs.

DISCUSSION

GENERALIZABILITY AND LIMITATIONS

With respect to the previously published empirical findings, it was found that the qualitative and quantitative evidence triangulated one another (Jick 1979). It was also found to be representative of the state of California and—to the extent that they were typical of national programs such as D.A.R.E. and LST—could be nationally generalized.

This research should be limited, though, by the fact that the interviewers were initially highly skilled in clinical forms of evaluation research. In addition to this, they received extensive additional training in the developed EOT. Because of human subjects considerations, such as fully informed consent, and potential EOT psychological impact, replication of this promising research might be undertaken only with considerable and proper training as described here, and as previously cited. Consequently, the extent that it is possible to skillfully apply these methods in qualitative and evaluation

research is the extent to which the methodology itself is worthy of further exploration.

Finally, although these results are promising, one is reminded that this procedure is exploratory. The results could be explained by chance or an additional factor such as the existing power dynamic between interviewer and respondent(s) or other unknown factors. These explanations are unlikely however, because of the consistency of the findings themselves and the absence of such "richer" data when the EOT was absent.

USE OF EOT MAY YIELD RICHER DATA

The evidence presented herein supports Mischler's theory regarding the importance of accounting for those "context-bound shades of meaning" regarding the interview situation. By training interviewers in what Brown (1996) referred to as "conscious self-involvement," the evidence seemed richer than is traditionally found in qualitative and evaluation research.

Specifically, when using EOT, respondents often accounted for human factors helping to explain the interviewer-respondent relationship. Many subsequently provided useful program implementation or effectiveness data. This apparent effect was called the "observation disclosure pattern." Conversely, in both youth and adult interviews, such a data pattern was largely absent when the EOT was absent.

Equally important, while using the EOT, several unexpected findings emerged. First, we discovered the depth of adult understanding of program ineffectiveness. In many cases, adults attributed program failure directly to influence methods used in programs such as D.A.R.E. and LST. As shown in the exemplar presented above, respondents believed that use of programmatic resources devoted to D.A.R.E. or LST was at the expense of other potentially more effective program resources, such as counseling services.

Second and most unfortunately, researchers observed cognitive dissonance among the youth who received the program, resulting directly from the drug education curriculum and its delivery processes. As the EOT was used, young people appeared more willing to forthrightly address conflicting issues and the precise nature of the conflict, such as the fear tactics used in drug education.

Finally, use of the EOT appeared to elicit substantial evidence regarding inconsistent zero-tolerance policy enforcement. This was evident in the exemplar presented above, in which on-campus cigarette smoking was acceptable for an adult school parent but not for a student who was 18 years old.

VALIDITY AND RELIABILITY WHILE USING THE EOT: FURTHER EXPLORATION IS MERITED

Although the results as described above remained consistent over the course of the study, the valid and reliable application of the EOT methodology in qualitative evaluation merits further exploration. This model may be best applied when the researcher wants to better test assertions while being directly involved in the interview situation. The extent that this method provides for a fuller understanding of the interviewer-respondent relationship, as well as evaluation phenomena, is the extent to which evaluation validity may be enhanced.

Equally important, Kirk and Miller (1986) pointed out that constant comparisons of data, as was conducted in this evaluation, are particularly sensitive to discrepancies between the researchers' interpretations and meanings actually intended by respondents. Because respondents are provided with opportunities to explore human factor dimensions during interviews, such comparisons using the EOT may provide an enhanced ability to locate the most significant evaluation phenomena, perhaps even how and why they are occurring. Because of the nature of the data collected then, EOT interviews may provide an enhanced ability to distinguish such interview effects from evaluation phenomena.

To a large extent, qualitative evaluation reliability depends on validity. Questions regarding reliability include but are not limited to, "Could these results be repeated if this study were administered in a similar fashion again" (Kirk and Miller 1986, 42)? Finding consistency in occurrences of phenomena about programs—such as youthful dissonance regarding D.A.R.E. or LST—often implies that the findings were initially valid.

In addition, the systematic yet dynamic aspect of the EOT approach is particularly salient because it serves as an in situ check on data—for consistency as well as depth. With respect to consistency, the distinction between the two different kinds of interview levels is worthy of note. The traditional interview schedule used in this study asked direct questions about the program and accompanying processes being evaluated. The EOT, on the other hand, addressed the nature of the interaction as Mischler described at the outset of this article. Probing of this kind is called "metacommunication," defined as "verbal and non-verbal communication that defines or clarifies the relationship between people who are in communication" (Brown 1996, 70). In this exploratory research, we focused primarily on nonverbal dimensions of the interaction as previously defined. The surprise discovery was the extent to which, as the team probed at the metacommunicative level regarding the

interaction (e.g., "you're smiling"), respondents linked these nonverbal cues regarding the interaction to evaluation phenomena. Consequently, the data available for constant comparison included consistent responses to the traditional semistructured interview questions as well as human factor probes. In a nontraditional sense, then, use of the EOT may offer an alternative form of triangulation (Jick 1979). In this case, the confirmatory evidence emerges from data collection on alternative levels (traditional interview questions vs. metacommunicative probes) rather than alternative sources (qualitative vs. quantitative).

With respect to the depth of evaluation findings in support of reliability, early on in the study, the research team confirmed outside research regarding drug education's inherent challenges. In addition, however, the evidence from the combination of the interview schedule and use of the EOT seemed to elicit a consistent and expanded understanding of challenges in such programs. Consider the example of youthful cognitive dissonance briefly mentioned earlier, as a report from the National Academy of Sciences noted in a review of our research:

> Brown and Kreft (1998) argue that the "no use" messages typically conveyed in universal prevention programs actually increase use among those most at risk for using. These youths are more knowledgeable about drugs and their effects than prevention curricula assume, and the naïve messages conveyed in the programs serve to create cognitive dissonance in the minds of these youths. (Manski, Pepper, and Petrie 2001, 218)

Within and between respondents, rather than emerging from traditional evaluation-focused questions, richer findings—such as the specific nature of cognitive dissonance caused by drug education—also emerged by addressing nonverbal human factors when using the EOT. If rigorous systematic observations are occurring as described in the article's methods—inquiring from the same respondent(s) using both evaluation-oriented questions as well as the EOT—it appears that the latter technique may offer significant checking for data reliability.

CONCLUSIONS

Based on the evidence contained herein, the EOT holds promise for qualitative and evaluation research. These assertions are supported:

- In collecting qualitative interview evaluation data, it is necessary but insufficient for the interviewer to pay attention to the content of spoken information.

During interview interaction, important human factors contextualizing the interaction and/or study phenomena cues present themselves. Many interviewers are not trained in methods to appropriately account for such nonverbal human factor cues. Based on the evidence, such training is important.

- Specialized interviewer training regarding human factors provides researchers with a valuable tool to increase data richness.

In closing, although more research is needed, the EOT helps researchers account for human factors during the qualitative interview, thus enhancing policy/program evaluation research methods.

REFERENCES

American Heritage Dictionaries, eds. 2000. *The American heritage dictionary.* 4th ed. Boston: Houghton Mifflin. http://dictionary.reference.com/search?q=non-verbal (accessed January 12, 2005).

Berelson, B. 1952. *Content analysis in communication research*. Glencoe, IL: Free Press.

Berg, D. N., and K. K. Smith, eds. 1988. *The self in social inquiry: Researching methods*. Newbury Park, CA: Sage.

Berger, P. L., and T. Luckmann. 1967. *The social construction of reality.* Garden City, NY: Anchor Books.

Blumer, H. 1969. *Symbolic interaction*. Englewood Cliffs, NJ: Prentice Hall.

Brown, J. H. 2001. Youth, drugs and resilience education. *Journal of Drug Education* 31 (1): 83-122.

Brown, J. H., and M. D'Emidio-Caston. 1995. On becoming at-risk through drug education: How symbolic policies and their practices affect students. *Evaluation Review* 19 (4): 451-92.

Brown, J. H., M. D'Emidio-Caston, and B. Benard. 2000. *Resilience education*. Thousand Oaks, CA: Sage.

Brown, J. H., M. D'Emidio-Caston, and J. Pollard. 1997. Students and substances: Social power in drug education. *Educational Evaluation and Policy Analysis* 19 (1): 65-82.

Brown, J. R. 1996. *The "i" in science: Training to use subjectivity in research.* Oslo, Norway: Scandinavian University Press.

California State Department of Education. 1992. *California basic educational data system*. Sacramento: California State Department of Education.

D'Emidio-Caston, M., and J. H. Brown. 1998. The other side of the story: Student narratives on the California Drug, Alcohol, Tobacco Education Program. *Evaluation Review* 22 (1): 95-117.

Festinger, L. 1957. *A theory of cognitive dissonance*. Palo Alto, CA: Stanford University Press.

Garfinkel, H. 1969. *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Geertz, C. 1973. "Thick description": Toward an interpretive theory of culture. In *The interpretation of cultures*, 3-30. New York: Basic Books.

Gilchrist, V. J. 1992. Key informant interviews. In *Doing qualitative research*, ed. B. F. Crabtree and M. Miller, 70-89. Newbury Park, CA: Sage.

Glaser, B. G., and A. L. Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.

Horowitz, J., and J. H. Brown. 1996. Confluent education and evaluation research. In *Advances in confluent education: Integrating consciousness for human change*, ed. J. H. Brown, 113-42. Greenwich, CT: JAI.

Jick, T. D. 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly* 24:602-61.

Kirk, J., and M. Miller. 1986. *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.

LeCompte, M. D., and J. Preissle. 1993. *Ethnography and qualitative design in educational research*. San Diego, CA: Academic Press.

Manski, C. F., J. V. Pepper, and C. V. Petrie, eds. 2001. *Informing America's policy on illegal drugs: What we don't know keeps hurting us*. Washington, DC: National Academy.

Marshall, C., and G. B. Rossman. 1989. *Designing qualitative research*. Newbury Park, CA: Sage.

Mischler, E. 1986. *Research interviewing*. London: Harvard University Press.

Sanders, J. R. 1994. *The program evaluation standards*. Thousand Oaks, CA: Sage.

Schutz, A., G. Walsh, and F. Lehnert. 1967. *The phenomenology of the social world*. Evanston, IL: University of Chicago Press.

Spradley, J. S. 1979. *The ethnographic interview*. New York: Holt, Rinehart & Winston.

Strauss, A. L., and J. Corbin. 1990. *Basics of qualitative research*. Newbury Park, CA: Sage.

Watzlawick, P., J. Weakland, and R. Fisch. 1974. *Change: Principles of problem formation and problem resolution*. New York: Norton.

Zelditch, M. 1962. Some methodological problems of field studies. *American Journal of Sociology* 62:566-76.

Joel H. Brown is a researcher/practitioner with more than 15 years of experience in mixed-methods evaluation and risk/resilience research in education and related systems. He integrates this research with policy/program development and training. In addition to numerous scholarly publications—including coediting a special issue of Evaluation Review*—he has published three books and is lead author of Corwin Press/Sage Publication's book titled* Resilience Education*. His work has been cited by the National Academy of Sciences, and he has given hundreds of speeches in a variety of settings. His commentary has been solicited in myriad media outlets, including ABC News 20/20, CNN, and* The New York Times*. Currently, he is a professor of education at the University of Oklahoma. He is also the executive director of the Center for Educational Research and Development (www.cerd.org), a not-for-profit organization committed to the development of resilience-based research, policies, and programs.*