

# Salvaging Quantitative Research With Qualitative Data

Donn Weinholtz  
Barbara Kacer  
Thomas Rocklin

*Through presentation of two case studies, this article illustrates just how ambiguous and misleading results from quantitative studies can be if not supplemented by qualitative data. The focus is on the salvaging power of qualitative methods and their ability to ensure some return on an investment that might otherwise be partially or completely lost.*

**I**t is no longer provocative to merely announce that quantitative and qualitative research methods can be fruitfully combined. The point has already been made repeatedly (e.g., Fielding & Fielding, 1986; Miles & Huberman, 1984; Reichardt & Cook, 1979). Furthermore, even those expressing strong misgivings about the "capture of qualitative inquiry by the quantitative approach," due to the accompanying blurring of important underlying objectivist and realist assumptions, acknowledge that "at the level of applying specific individual procedures...the two approaches can be mixed" (Smith & Heshius, 1986, p. 4). Consequently, we will not argue here over whether quantitative and qualitative approaches can be combined. Rather, via presentation of two case studies we will illustrate just how ambiguous and misleading results from quantitative studies can be if not supplemented by qualitative data. In doing so, we will address an issue generally ignored in the literature and also show how use of sup-

plemental qualitative methods by quantitative researchers can serve as a prudent hedge against obtaining inconsequential or erroneous results.

The first case that we will discuss illustrates how qualitative data can permit insightful interpretation of studies that yield no significant quantitative findings. The second will demonstrate how qualitative data can permit detection of errors in quantitative data analysis procedures by providing important reference points for numeric values. In both cases, even though substantial care was taken by the researchers in their quantitative designs, the qualitative data played an important role in rendering the studies useful. Thus we will focus on the "salvaging power" of qualitative methods in reference to their ability to ensure some return on an investment that might otherwise be partially or completely lost.

#### A STUDY OF SMALL GROUP INSTRUCTION OF COMPUTER APPLICATIONS

The first case, previously reported by Kacer, Weinholtz, and Rocklin (1991), investigated the impact of small group instruction on attitude and achievement of students learning computer applications (e.g., word processing, spreadsheets, and databases). Forty-nine teacher education students (44 females, 5 males) were randomly assigned to either an individual ( $n = 15$ ) or a group ( $n = 17$  dyads) condition. Using the Appleworks program to learn the three designated computer applications skills, subjects met in their individual or group conditions for 3 consecutive weeks to complete an instructional assignment for each of the three applications. (On these assignments, subjects within the dyadic condition each received the same score for their performance.) Also, subjects from both conditions were individually administered a comprehensive postinstructional test, involving creating new files and manipulating existing files. Finally, before and after instruction, all subjects completed a questionnaire measuring attitude toward instruction.

In addition to the numerically scored and analyzed assignments, tests, and questionnaires, the researchers collected two types of qualitative data. First, subjects in both the individual and group conditions were audiotaped while working at their computers on their instructional assignments. Second, at the conclusion of the study, subjects from each condition (4 individuals plus 9 members from dyads) were randomly selected for debriefing interviews. Following completion of the study's quantitative data analysis, these qualitative data were analyzed via the review and categorical coding of transcribed audiotapes and interview notes, as prescribed by Lofland (1971).

To briefly summarize the results of the quantitative analysis, neither multivariate nor univariate analysis of variance procedures re-

vealed significant differences between individuals and groups on either the assignment, test, or attitude measures. The researchers found the results encouraging, in that dyadic computer use made more efficient use of equipment without incurring any penalty in learning. But they were also disappointed. Based on personal experience and an extensive review of the small group instruction literature, it had been assumed that distinctive differences between individuals and dyads would exist, but the carefully developed quantitative comparison provided no indication of any difference, nor any clue as to where differences, if they existed, might be found.

Faced with this situation, the traditional quantitative researcher is forced to review the research effort in an attempt to either explain the present findings or design future research. The researcher asks, "Did we have adequate statistical power?" "Were our measures adequately reliable and valid?" "Was our treatment powerful enough?" "Was the treatment implemented as planned?" "Did subjects perceive stimuli (or understand questions, or react to manipulations) as we expected?" Some of these questions can be answered rather directly (e.g., statistical power, reliability of measures). But these are the very questions that can and should be answered before the research is conducted. Other questions (e.g., treatment integrity, subjects' reactions) cannot be answered *a priori*. Neither can they necessarily be answered from any of the data collected in the quantitative research. Worst of all, the specific form of some of these questions cannot be anticipated before the data are collected.

Qualitative data can play a particularly useful salvaging role both in understanding the findings and in suggesting future research. In the small group instruction/computer applications study, this certainly proved to be the case. Analysis of 11 randomly selected, transcribed audiotapes revealed that students working in the individual condition asked three times the number of questions of available teaching assistants as did students working in dyads. This finding indicated that rather than establishing a pure comparison between individual and group performance and attitudes, the study compared group performance with that of a condition that was something akin to a tutor-tutee relationship. Furthermore, conversations within dyads consistently indicated delayed use of the keyboard and the initiation of planning activities, whereas students in the individual condition regularly turned directly to use of the keyboard and asked questions of teaching assistants that indicated little in the way of planning activities. Finally, dyads asked a greater percentage of conceptual-level questions (45% to 33%) than did individuals, and the questions raised by dyads consistently indicated conceptual understanding, whereas those from individuals revealed bewilderment.

Analysis of the interview data also revealed differences between the two conditions. Those from the individual condition all indicated that the classroom experience had been enjoyable and that instruction had been helpful. Beyond these points, no clear opinion patterns emerged. However, interviews with dyad members offered additional information. For example, learning from one's partner, relatively independent of the instructor, was consistently reported. Also, feedback from one's partner was valued, as was the knowledge that one's partner was putting forth a conscientious effort. As one student put it, "I would rather do my own work and make sure it's done correctly than be with somebody else who doesn't do it and gets the same grade that I do."

In combination, the audiotape data and interview data did not change the "no-difference" quantitative findings, but they did make the numeric results more understandable and interesting. They made it clear that the study's individual condition had unexpectedly metamorphosed into a modified group condition. Also, they provided evidence that the study's outcome measures were insensitive to an apparent higher level learning outcome within the dyad condition. Finally, they revealed a condition (lack of perceived mutual effort) that might undermine group performance.

These qualitative findings helped to establish connections with previous research, such as Webb's (1984) finding that individuals and small groups working on computer tasks (although showing no achievement differences) adopted different learning strategies. Even more important, they suggested enticing directions for future research. For example, the individual condition might be more carefully controlled and additional higher order learning measures might be incorporated. Also, interactions and attributions of lower performance dyads might be analyzed to further assess the impact of perceived mutual effort on group performance. From such analyses, recommendations for enhancing mutual effort might be developed and tested.

All of these possibilities were truly exciting. They provided insight and momentum to a research initiative that otherwise would have stalled due to lack of plausible explanations for the outcomes of the original study. Using a term made popular by Parlette and Hamilton (1972), the qualitative findings "illuminated" the quantitative results, thereby decreasing the need for the researchers to offer shot-in-the-dark interpretations about what occurred during the study

#### A STUDY OF TEACHING BY UNIVERSITY HOSPITAL ATTENDING PHYSICIANS

The second case actually involved two studies, one correlational the other experimental, linked together in a single effort to modify

and improve the teaching of university hospital attending physicians (clinical faculty) in the small group teaching and working setting known as attending rounds. Attending rounds occupy approximately 2 hours per day and involve several medical students, first-year resident physicians, a senior resident physician, and the presiding attending physician in a rich, experiential teaching and learning activity.

Briefly, the correlational study (Weinholtz, Everett, Albanese, & Shymansky, 1986) involved collection of behavioral data on the attending round teaching of 41 attending physicians.

Each physician was observed for 8 to 10 hours (1 week's attending rounds), and teaching behaviors were recorded by trained observers carrying portable computers programmed to accept data from the Attending Round Observation System (Weinholtz, Albanese, Zeitler, Everett, & Shymansky, 1986), a behavioral recording scheme specifically designed for use in the setting under study. Following each observation period, the medical students and residents on each attending physician's team were asked to rate specific aspects of the attending physician's teaching as well as his or her overall teaching performance. Students also completed two open-ended items regarding the attending physician's teaching, and observers scored brief qualitative diaries regarding each day's activities. Correlations were calculated between percentages of time that attending physicians allocated to particular teaching activities and to the learners' ratings of teaching effectiveness.

In addition to yielding correlational findings indicative of teaching behaviors that learners perceived to be effective, this study provided baseline teaching behavior data that were used as the premeasures in the follow-up experimental study designed to assess whether one-shot instructional consultations could yield changes in attending physician teaching behaviors and ratings by learners of teaching effectiveness (Weinholtz, Albanese, Zeitler, & Everett, 1989). Fourteen of the 28 physicians who were able to continue into the second phase were randomly assigned to a treatment condition and 14 were assigned to a control group. Those in the treatment group received a consultation 1 to 3 weeks prior to a second round of observations and ratings. The consultations provided the attending physicians with feedback on their prior behavioral data, ratings data, and on the qualitative comments from learners and observers. This feedback was used to negotiate a set of target areas in which the attending physicians might experiment with new teaching approaches.

The 14 attending physicians assigned to the control group also received consultations, but not until after the second round of observation and ratings. In addition to the behavioral and ratings data collected as out-

come measures, all attending physicians participating in the study completed a brief evaluation form focusing on the observation and consultation process.

The correlational study identified 18 teaching behaviors, occurring across various attending round contexts and settings, that were pertinent for use in the consultations. The experimental study revealed that the consultations did provoke behavioral change, but that increase in learner ratings of teaching was not significantly different from a similar increase in the control group ratings. Also, the attending physicians' evaluations of the observation and consultation process were highly positive. In combination, these two quantitative studies were quite successful in making a contribution to a relatively uncharted area of educational research. They were published and are now frequently cited in the clinical, medical education literature. So, what is the issue here?

The issue is that, if not for the collection of a modest amount of qualitative data (the learner responses to two open-ended questions on the teacher rating form and the brief diaries maintained by the observers), the study would have foundered due to an undetected and potentially catastrophic data analysis error. To make a long and painful story relatively short, the initial correlational analysis revealed some fairly straightforward and also some relatively "counterintuitive" findings. Based on these findings, consultations for the experimental phase were to be devised. However, while designing the consultations the learner comments and observer diaries (which were added to the study almost as an afterthought) had to be consulted. While developing the first consultation, a striking problem became apparent. The physician's behavioral profile was grossly inconsistent with the qualitative comments and accounts. A quick check revealed that this was also the case for the data on several other physicians. Somewhere, something had gone wrong, but it was not clear where.

Eventually, the source of the error was found. Approximately halfway through the correlational study, a mistake had been made in the assignment of identification numbers to the attending physicians. Down the line, there was a resulting mismatch of behavioral and ratings data through half of the two data sets. This resulted in one physician's behavioral data being correlated with another's ratings data. (No wonder there were counterintuitive findings!) The entire correlational analysis had to be repeated.

Fortunately for all involved, the data for the attending physician scheduled for the first consultation just happened to have come from the second half of the data sets where the error existed. This was a stroke of luck, but even with that luck the discovery, made in the waning hours before the first consultation was supposed to have been delivered, would never have occurred without the col-

lection of the small amount of qualitative data that was designed into the studies. Without that data, consultations would have been offered and recommendations made based on an error-filled analysis. Half of the physicians would have been offered supposedly "hard," "numeric" observational profiles of their (actually others') teaching behaviors. We can only guess how they would have reacted. Some might have scratched their heads and said, "Gee, I didn't realize that I did that," but some (quite rightfully) would probably have announced that their profiles were preposterous because they never engaged in some of the behaviors presented to them.

Obviously, the error described here could have been detected by more carefully grooming the data. However, many precautions had been taken. The researchers involved thought they had monitored the data as carefully as possible. It was one research assistant's half-time job just to tend to the data sets, and that individual's efforts were double-checked by one of the researchers. What happened (as can happen so often) was one minor human error that occurred while navigating in a sea of numbers, an error not detected by either of two individuals. In this downwardly spiraling scenario, a relatively small amount of qualitative data (without prior intent on the part of the researchers) served as a failsafe mechanism preventing a horrible waste of invested time and energy. Although the experimental study timetable was thrown off schedule by having to redo the correlational analysis, both studies were salvaged because the qualitative data cast the numbers in verbal description that made the error much more readily detectable. We can only wonder how many other studies have suffered irreparable damage from not having a qualitative safety net in place. There is no way to gauge the extent of this sort of problem because researchers do not often report such embarrassments. However, even if the problem is infrequent, the cost is too great to those involved not to take precautions against such consequences.

#### FURTHER LESSONS

Beyond the points already made, what lessons can be derived from these two cases? First, the collection of qualitative data was neither a particularly obtrusive nor labor-intensive effort. The audiotapes in the small group/computer study were the only data recorded simultaneously with quantitative data collection, and there is no evidence that the tape recorder altered individual or group responses to the assignments. The debriefing interviews at the end of this study were the single most time-consuming effort, but the use of a random sample of subjects for interviewing made the task manageable (while admittedly sacrificing some data).

Second, analysis of the data required additional time and effort, especially in the small group/computer study. However, the insights gained in both studies clearly warranted the additional investment, which was quite small compared to the overall effort required for mounting and executing their quantitative components.

Third, both studies had a small group focus, and qualitative research methods are particularly suited for highlighting group processes not easily captured by quantitative measures. Other researchers conducting quantitative studies involving small group conditions would seem particularly well advised to incorporate supplemental qualitative methods. Nevertheless, even those not investigating small group phenomena could benefit substantially from doing so, particularly by making efforts to ensure that their experimental treatments have had their expected impact. For such purposes, debriefing interviews can be particularly effective.

To illustrate, a doctoral student on whose dissertation committee one of us was once asked to sit completed an elegant multivariate analysis of variance in a study comparing group performances on speeded achievement tests (Conboy, 1986). The critical variable in the study was the imposition of awareness levels of target goals for the different groups. It was assumed that creation of high awareness levels (particularly of initial failure) would enhance motivation (and consequently performance). However, the extraordinarily careful analysis revealed little other than an obscure interaction. On the other hand, a "manipulation check" via a short debriefing interview conducted by the researcher (again almost as an afterthought) revealed that "fewer than 20% of the subjects could identify their assigned goals on difficult problems, and only 35% could do so on easy problems." This poststudy assess-

ment provided ample evidence that the expected treatment never really took hold, a useful finding that might have yielded even more valuable information if more in-depth, probing interviews of subjects had been conducted following completion of the quantitative data collection. How many times is the impact of a treatment rejected when the treatment has not really been established? We simply do not know, but it is plausible that in most experimental studies debriefing interviews in combination with unobtrusive observations could determine if a "take" occurs.

Conducting research of any kind is a high-stakes endeavor. It is not just one's intellectual acumen that is tested; one's career and academic credibility are held in the brink. It simply does not make sense to conduct "black box" studies that yield bewildering results. For quantitative researchers, the judicious use of qualitative methods offers an efficient and powerful tool for self-correcting studies gone astray. Much good work can be salvaged through forethought of design.

## REFERENCES

- Conboy, J. E. (1986). *Goal awareness and self-efficacy as determinants of the constructive effects of failure*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Fielding, N. G., & Fielding, J. L. (1986). *Linking data*. Beverly Hills, CA: Sage.
- Kacer, B., Weinholtz, D., & Rocklin, T. (1991). The impact of small group instruction upon attitude and achievement of students' learning computer applications. *Computers in the Schools*, 8, 357-360.
- Lofland, J. (1971). *Analyzing social settings*. Belmont, CA: Wadsworth.
- Miles, M., & Huberman, M. (1984). *Qualitative data analysis*. Beverly Hills, CA: Sage.
- Parlette, M., & Hamilton, D. (1972). *Evaluation as illumination: A new approach to the study of innovative programmes*. Occasional Paper

9, Center for Research in the Educational Sciences, University of Edinburgh.

- Reichardt, C., & Cook, T. (1979). Beyond qualitative versus quantitative methods. In T. Cook & C. Reichardt (Eds.), *Qualitative and quantitative wetlands in evaluation research* (pp. 7-32). Beverly Hills, CA: Sage.
- Smith, J. K., & Heshius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15(1), 4-12.
- Webb, N. M. (1984). Microcomputer learning in small groups: Cognitive requirements and group processes. *Journal of Educational Psychology*, 76, 1076-1088.
- Weinholtz, D., Albanese, M., Zeitler, R., & Everett, G. (1989). Effects of individualized observation with feedback on attending physician clinical teaching. *Teaching and Learning in Medicine*, 1(3), 128-134.
- Weinholtz, D., Everett, G., Albanese, M., & Shymansky, J. (1986). The attending round observation system: A procedure for describing teaching during attending rounds. *Evaluation & the Health Professions*, 9, 75-89.
- Weinholtz, D., Albanese, M., Zeitler, R., Everett, G., & Shymansky, J. (1986). Effective attending physician teaching: The correlation of observed instructional activities and learner ratings of teaching effectiveness. In *Proceedings of the Twenty-fifth Annual Research in Medical Education Conference* (pp. 273-278). Washington, DC: Association of American Medical Colleges.

---

*Donn Weinholtz, Ph.D., is the dean of the College of Education, Nursing and Health Professions at the University of Hartford.*

*Barbara Kacer, Ph.D., is an associate professor of education at Western Kentucky University.*

*Thomas Rocklin, Ph.D., is a professor of education at the University of Iowa.*