Principles and Practices: An Empirical Examination of Qualitative Research in the Journal of Counseling Psychology

William T. Hoyt and Kuldhir S. Bhati
University of Wisconsin—Madison

This article examines the 50 qualitative studies published in the Journal of Counseling Psychology (JCP) over a 15-year period in light of methodological principles advocated by qualitative theorists. The match between practices and principles is not high. In the modal investigation, coders (most of whom did not interact with or observe participants) worked from transcripts of a 60-min interview conducted in a setting convenient for the researcher. Researchers endorsed the need to bracket their own subjective experiences and used auditors to enhance reproducibility of findings. Trend analyses suggest that qualitative research in JCP has tended to drift further from qualitative principles over time. The authors consider the implications of these findings for the potential of qualitative methods to inform science and practice in counseling psychology.

Keywords: qualitative methods, philosophy of science, research practices, review

Counseling psychology has a relatively long history of receptivity to qualitative methods, with calls for expansion of the methodological repertoire dating back at least 30 years (e.g., Gelso, 1979; Goldman, 1976). Three main reasons have been advanced for the compatibility of qualitative or phenomenological inquiry with research goals of counseling psychologists. First, some advocates of qualitative methods (e.g., Hoshmand, 1989) have asserted that quantitative methods are mired in a positivistic philosophy of science that is not suitable for studying some aspects of human experience, such as personal meanings associated with life events. Second, phenomenological inquiry can be conducted with small samples and provides rich idiographic descriptions of participant experiences. These methods are, therefore, likely to be attractive to practitioners, both as researchers and consumers of research (Hoshmand & Polkinghorne, 1992), thus advancing the ideal of the scientist–practitioner model (Heppner et al., 1992). Third, phenomenological or ethnographic inquiry may be especially relevant to the study of rare or rarely researched populations.

The ongoing process of legitimization of qualitative methods in counseling psychology has been described in dramatic language as a “paradigm shift” (Ponterotto, 2005a, p. 126), a “tectonic change” (O’Neill, 2002, p. 196), and a “fifth force” in psychology (Ponterotto, 2002, p. 394) that will weaken the hegemony of quantitative approaches and enhance openness to alternative paradigms for acquisition of knowledge in the discipline. Yet, despite evident enthusiasm for qualitative approaches, examinations of published research in psychological journals generally (Rennie, Watson, & Monteiro, 2002) and in journals published by the APA (Kidd, 2002) show slow progress in dissemination of qualitative findings through conventional research outlets. On the basis of a content
analysis of journal editorial statements and interviews with editors from 10 APA journals, Kidd (2002) concluded that one barrier to publication of qualitative work in these highly visible journals is precisely their tradition of equating scientific research with quantitative methods and the consequent perceived and actual ambivalence on the part of editors about the potential contribution to knowledge of qualitative scholarship. Specifically, editors raised concerns about lack of generalizability and about potential for researcher biases to compromise validity of qualitative findings, and 4 of the 10 editors interviewed indicated that they did not believe that qualitative work was appropriate for their journals.

Qualitative methodologists have shown some willingness to adapt their language and practices to address challenges from quantitatively trained researchers. Perhaps the best known example is Guba and Lincoln’s (1994; cf. Guba, 1981) formulation of criteria for evaluating the trustworthiness of qualitative (or naturalistic) inquiry. Guba (1981) proposed four considerations (credibility, transferability, dependability, and confirmability) that should be addressed in evaluating qualitative research. These were labeled as parallel criteria, or even as extrinsic criteria (Lincoln, 1995), in acknowledgment of the fact that they addressed aspects of validity (viz., internal validity, external validity, reliability, and objectivity) that evolved in the quantitative research tradition. These parallel criteria have been embraced by the qualitative research community and form the core of many recommendations for the evaluation of the methodological quality of qualitative scholarship (e.g., Elliott, Fischer, & Rennie, 1999; Guba & Lincoln, 1994; Morrow, 2005).

This willingness to submit qualitative research to the standards of the quantitative scientific mainstream is understandable. For qualitative researchers, earning credibility in the eyes of the quantitative establishment is desirable for both professional standing and dissemination of scholarly work. Adherence to extrinsic criteria may become problematic, however, if it changes the character of qualitative work, creating a disjunction between the intrinsic principles that underlie naturalistic inquiry and the practice of qualitative researchers in the field. This potential problem is exacerbated by the fact that Guba’s (1981) quasi-quantitative criteria are to some extent at odds, both logically and pragmatically, with authentic principles and practices underlying qualitative research. To show why this is so, we present a brief overview of the different cultures of inquiry embraced by qualitative and quantitative researchers.

QUAL and QUAN: Differing Research Cultures

It may seem odd to characterize research practices as cultures. Scientific methods traditionally have been viewed as rationally derived techniques for discovering the structure of the natural world (e.g., Hempel, 1952/1970). However, Kuhn (1962) famously observed that existing practices (or paradigms) within a scientific discipline are never fully justifiable either rationally or empirically. Indeed, as Gergen (2002) noted, a primary focus of the postmodern critique of positivism and postpositivism has been to challenge the received view that these methodologies are the only source of valid knowledge and to “draw attention to the character of scientific psychology as a particular kind of cultural tradition” (p. 463). In this section, we elaborate on this notion of scientific traditions as cultures by identifying some of the defining features of quantitative (QUAN) and qualitative (QUAL) research cultures.

Defining Features of QUAL and QUAN Research Cultures

Culture may be broadly defined as a socially transmitted collection of knowledge, habits, and skills (de Waal, 2001). Important aspects of scientific cultures (akin to what Kuhn, 1962, called scientific paradigms) are assumptions about the nature of scientific knowledge and shared exemplars representing prototypical knowledge-generating activities. Morrow and Smith (2000) described three main areas in which QUAL methods diverge from the dominant QUAN research culture: (a) focus of inquiry, (b) research setting, and (c) role of the researcher.

1 Several reviewers noted that the QUAL principles articulated here apply to phenomenological and interpretive approaches (and probably also to critical theory, although this paradigm is less frequently represented in the studies we reviewed) but less well to approaches such as consensus qualitative research (CQR; Hill, Thompson, & Williams, 1997). This raises the question of whether the principles on which these 50 investigations are evaluated in our study are sufficiently inclusive.

Focus of inquiry. Morrow and Smith (2000) use the terms idiographic and nomothetic (Allport, 1937) to describe the different research foci of QUAL and QUAN cultures. In QUAL research, the focus is on a rich and complex exploration of the experiences of a small number of individuals. In this idiographic approach, the goal is to attain a deep understanding of one or more people who are regarded as unique in their own right. By contrast, QUAN researchers adopt a nomothetic perspective, in which participants’ representativeness (of the population about which one seeks generalizable knowledge) rather than their uniqueness is salient. For a nomothetic inquiry, the focus is on understanding not the individuals themselves but a small set of characteristics or traits (i.e., variables) that these individuals possess in varying degrees. QUAL researchers seek to refute or support hypotheses about causal associations among variables—and ultimately to uncover general laws of behavior that apply to large groups (populations).

Research setting. Because of the idiographic focus in QUAL work, investigators in this tradition wish to establish a considerable acquaintance with their research participants. Thus, they spend a substantial amount of time with each participant, ideally in that person’s typical environment. This emphasis on field settings flows from an appreciation of the importance of the situation (context) in determining both behavior and experience. Thus, “the
researcher must be present in participants’ natural worlds and everyday lives and close enough, spatially and psychologically, that participants will reveal the meaning they make of their experiences” (Morrow & Smith, 2000, p. 201). QUAN researchers also recognize the importance of situation or context as a determinant of behavior, but they see this as a source of nuisance variance to be controlled or substantive variance to be standardized (i.e., manipulated or measured) and studied in its own right. Therefore, the prototypical setting for QUAN research is the laboratory, where contextual variables are (in principle) under the control of the investigator. When QUAN research is conducted in field settings (e.g., schools, clinics), investigators may still do their best to control contextual variables by standardizing testing conditions, manualizing interviews or treatments, and carefully training experimenters or treatment staff.

**Researcher role.** QUAL researchers reject a rigid dichotomy between facts and values. No indisputable facts are available to psychological researchers because all observation is conditioned to some degree on perceptual and judgmental processes of the observer. Thus, the QUAL research culture embraces the role of researcher as instrument, in a frank acknowledgment that the conclusions and even the raw data of a QUAL investigation are jointly determined by the subjectivities of researchers and participants. Rather than trying to erase or ignore his or her own assumptions and biases, the QUAL researcher is encouraged to reflexively explore both preconceptions and evolving understandings in the research narrative.

By contrast, QUAN researchers in the positivist tradition attempted to delineate observed behaviors as a psychological domain that could be studied with a rigor comparable to the physical sciences (i.e., without distortion due to researcher subjectivity). This movement redefined psychology from the *science of mind* to the *science of behavior* (Leahey, 1992, pp. 269–270), a conception that has endured even after radical behaviorism fell from favor during the cognitive revolution of the 1950s and 1960s (Mahoney, 1991, chap. 4). As psychologists returned to the study of subjective (i.e., unobservable) experiences, positivism became untenable as a philosophical foundation for psychological inquiry and was replaced by postpositivist notions of an underlying reality that is latent (i.e., not directly observable or measurable), so that both measures and theories were to be validated by a process of successive approximations (Cronbach & Meehl, 1955), with attention to sources of error and bias in quantitative measures (Campbell & Fiske, 1959) and careful consideration and gradual elimination of plausible rival explanations for study findings (Campbell & Stanley, 1963).

In the postpositivist QUAN research culture, there is a healthy respect for the subtle mechanisms by which researchers may intentionally or unintentionally influence the outcome of experimental and nonexperimental investigations (Rosenthal, 2000; Shadish, Cook, & Campbell, 2002, chap. 3; Wampold, 2001, chap. 7). Thus, the role of the researcher is to be aware of the possible biases or expectancies on study findings and to minimize this potential confound by thoughtful design of the research project, which often involves distancing himself or herself from research participants. Participants’ primary contact would then be with research assistants who, ideally, would be naive to the researcher’s hypotheses.

Thus, both QUAL and QUAN research cultures accept the fundamental intersubjectivity of the research enterprise—the premise that participant responses do not reflect objective data but, instead, are at least partly conditioned on the person of the researcher. Both cultures encourage researcher awareness of this potential impact, but the implications for researcher role are very different in the two cultures. As Morrow and Smith (2000) summarize, “Whereas conventional [QUAN] researchers use this awareness to ‘bracket’ or attempt to eliminate subjective influences on the investigation, constructivists [i.e., QUAL researchers] hope to identify, for themselves and their audience, the ‘lens’ through which the research is conducted” (p. 201).

**Summary**

The nature of the research enterprise is quite different in QUAL and QUAN cultures, and each culture has evolved supporting institutions and practices for pursuing its ideal or paradigmatic activities. For QUAN researchers, these cultural traditions include a language for speaking about research quality (focusing on evaluative criteria, such as internal validity, external validity, construct validity, and statistical conclusion validity), a recommended set of design principles for reducing threats to validity (e.g., Kirk, 1995; Shadish et al., 2002), and criteria for preparation and review of research articles that allow readers to make informed judgments about the validity of inferences based on research findings.

The QUAL culture in psychology is newer, but qualitative work has been firmly established in related fields such as education, social work, medicine, sociology, and anthropology (Denzin & Lincoln, 2000; Kidd, 2002; Rennie, 1999). The QUAL culture in psychology can access a rich heritage of core practices in these disciplines, which have a far longer history of conducting, evaluating, and accepting qualitative research (Tashakkori & Teddlie, 1998). Although there are many broad general guidelines for qualitative researchers, the diversity of qualitative methods calls into question the existence of a single unified paradigm for this work (Morrow, 2005). However, Morrow and Smith (2000) propose the idiographic approach as one that most (and perhaps all) QUAL researchers would endorse, and they sketch clear contrasts with the nomothetic orientation of work in the QUAN tradition.

Thus, the two research cultures diverge in fundamental ways in both principles and practices. In this study, we set out to characterize the considerable body of QUAL work published in *JCP* over a 15-year period, focusing on whether the practices in these published studies conform to the principles just described.

**Research Questions**

To investigate the extent to which qualitative investigations published in *JCP* remain close to the principles of the QUAL research culture, we coded articles on the three dimensions identified above: (a) focus of research (nomothetic vs. idiographic); (b) setting (naturalistic vs. laboratory); and (c) researcher role (reflexive examination of intersubjectivity vs. attempts to bracket or otherwise banish the researcher’s subjectivity from the study). For each of these dimensions, and for one additional dimension of presentation of study findings, we identified one or more relevant
study qualities and present descriptive data on these qualities for 50 qualitative or mixed-method studies published in *JCP* over a 15-year period. A detailed description of study qualities coded is provided in the Method section.

**Method**

**Literature Search**

We used Ponterotto’s (2005b) classification of studies in *JCP* covering the 15-year period from 1989 to 2003. Ponterotto identified 39 QUAL studies and an additional 10 studies classified as mixed-method inquiries, meaning that they had a substantial QUAL component and also included some QUAN analyses. One article (Swagler & Ellis, 2003) included 3 separate studies, 2 of which used predominantly QUAL methods. Because these are separate studies involving distinct samples and methods, we treated them as 2 independent studies in our analyses. Thus, we coded 50 studies (39 considered to be pure QUAL studies by Ponterotto and 11 that he coded as mixed).

**Study Coding**

We independently coded each study for 12 attributes relevant to the three dimensions described above (focus of research, setting, and researcher role) and a fourth dimension (presentation of findings) to examine the extent to which findings were communicated in a manner congruent with the goals of QUAL research. These attributes are listed in Table 1, along with the kappa coefficient indexing initial interjudge agreement for the 11 categorical attributes and the intraclass correlation indexing interjudge reliability for the single quantitative attribute (time spent with each participant).

Kappa (Cohen, 1960) is a measure of the proportion of agreement corrected for chance. Fleiss (1981) suggests that kappa coefficients less than .45 represent poor agreement, whereas kappas greater than .75 represent strong agreement. Intermediate values of kappa (between .45 and .75) represent moderate interjudge agreement.

We computed the intraclass correlation coefficient (ICC) designated as ICC(3, 2) by Shrout and Fleiss (1979). This is the appropriate interjudge reliability coefficient when scores will be aggregated over two raters ($n_r = 2$), and all targets (in this case, all studies) will be evaluated by the same two raters (so that constant rater effects do not contribute to error variance in scores—see Hoyt & Melby, 1999, p. 339). ICCs greater than .70 are considered adequate evidence of reliability (Heppner, Kivlighan, & Wampold, 1999).

In general, agreement between raters was adequate or strong. Two study qualities (data available to all coders and setting for interactions or observations) had low kappas despite relatively strong percentages of agreement because of low base rates for most categories in this sample of studies. (Low base rates for categorical data are equivalent to restriction of range for continuous variables and tend to attenuate coefficients of agreement; Fleiss, 1981.) For one other study characteristic (proportion of coders having contact with participants), initial poor agreement reflected an ambiguity in the coding system. One of us initially interpreted this as the proportion of study authors rather than coders. Once this ambiguity was clarified, disagreements for this characteristic were easily resolved. For all study qualities, initial disagreements were resolved in discussion, with careful rereading of the relevant study descriptions.

**Focus of research.** Researchers engaged in idiographic inquiry emphasize a high level of acquaintance with participants and a collaborative relationship to invite participants to assist the researcher in understanding their subjective experience. In most QUAL studies published in *JCP*, data analysis has been conducted by a group of judges or coders, only some of whom have direct contact with (or opportunity to observe) participants. To gauge the degree of acquaintance between coders and research participants in these studies, we coded the following characteristics: (a) contact between any coder and participants (yes or no), (b) nature of contact (none, telephone, or face-to-face contact), (c) proportion of coders who have such contact with at least one participant, (d) data available to all coders (transcript or other written, audiotape, videotape, direct observation or interaction), and (e) average time spent with each participant.

**Setting.** Because it is desirable for QUAL researchers to have a sense of the lived world of their research participants, we coded studies according to whether interactions with or observation of participants took place in a naturalistic setting, defined as a real-world setting in which the behaviors of interest might normally be expected to occur. (Examples of nonnaturalistic settings included telephone interviews and in-person interviews conducted in a location convenient for the researcher.)

**Researcher role.** To quantify the extent to which the authors of these studies embraced the role of researcher as instrument, we examined study Method, Results, and Discussion sections to tabulate (a) use of auditors as a check on the findings of the main coders, (b) written reflection on the authors’ biases in the published article, and (c) explicit discussion of the desirability of minimizing or eliminating the effects of researcher bias.

**Reporting findings.** A final set of qualities examined the way that authors organized the published research report to communicate the rich, complex understandings sought by idiographic researchers. Characteristics under this heading included (a) presentation of free-standing quotations (at least one sentence in length) from participants, (b) presentation of extended quotations (at least five lines in length), and (c) inclusion of identifiers with the quotations. The last category has been recommended (Elliott et al., 1999) as a means to assist the reader in connecting quotations from the same source that may appear in different sections of the article, thus enhancing the sense of the participant as a multifaceted individual.

**Results**

Characteristics of the 50 QUAL and mixed QUAL–QUAN studies published in *JCP* over the 15-year period from 1989 to 2003 are shown in Table 1. Coded variables are grouped according to four important dimensions of difference in the QUAL and QUAN research cultures: research focus, research setting, researcher role, and reporting of findings.

In an initial set of analyses, we used the chi-squared test to assess differences in category frequency between mixed studies ($k = 11$) and pure QUAL studies ($k = 39$) for all 11 categorical
## Table 1

**Frequencies and Initial Agreement Indices for 12 Coding Dimensions**

<table>
<thead>
<tr>
<th>Category and code</th>
<th>All studies</th>
<th>Early studies</th>
<th>Late studies</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus of research (initial agreement coefficient)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do any coders have contact with participants? (( \kappa = .95 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>14</td>
<td>28</td>
<td>8</td>
<td>44</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>36</td>
<td>72</td>
<td>10</td>
<td>56</td>
<td>26</td>
<td>81</td>
</tr>
<tr>
<td>What type of contact? (( \kappa = .83 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = None</td>
<td>14</td>
<td>28</td>
<td>8</td>
<td>44</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>1 = Phone</td>
<td>14</td>
<td>28</td>
<td>2</td>
<td>11</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>2 = Face to face</td>
<td>22</td>
<td>44</td>
<td>8</td>
<td>44</td>
<td>14</td>
<td>44</td>
</tr>
<tr>
<td>What proportion of coders has such contact? (( \kappa = .03 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = Half or fewer</td>
<td>41</td>
<td>82</td>
<td>16</td>
<td>89</td>
<td>25</td>
<td>78</td>
</tr>
<tr>
<td>1 = More than half</td>
<td>9</td>
<td>18</td>
<td>2</td>
<td>11</td>
<td>14</td>
<td>44</td>
</tr>
<tr>
<td>Data available to all coders (( \kappa = .47 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = Written</td>
<td>44</td>
<td>88</td>
<td>14</td>
<td>78</td>
<td>30</td>
<td>94</td>
</tr>
<tr>
<td>2 = Audiotape</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3 = Videotape</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4 = Live observation</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average time spent with each participant [ICC(3, 2) = .88]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (min)</td>
<td>60</td>
<td>0</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Setting (initial agreement coefficient)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did any interactions or observations take place in a naturalistic setting? (( \kappa = .50 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>40</td>
<td>80</td>
<td>12</td>
<td>67</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>10</td>
<td>20</td>
<td>6</td>
<td>33</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td><strong>Researcher role (initial agreement coefficient)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were the main coders’ findings checked by independent auditors? (( \kappa = .82 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>16</td>
<td>32</td>
<td>10</td>
<td>56</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>34</td>
<td>68</td>
<td>8</td>
<td>44</td>
<td>26</td>
<td>81</td>
</tr>
<tr>
<td>Do the authors reflect on their own biases or preconceptions in the published article? (( \kappa = .72 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>21</td>
<td>42</td>
<td>10</td>
<td>56</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>29</td>
<td>58</td>
<td>8</td>
<td>44</td>
<td>21</td>
<td>66</td>
</tr>
<tr>
<td>Do the authors explicitly discuss the desirability of minimizing or eliminating these biases? (( \kappa = .53 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>26</td>
<td>52</td>
<td>15</td>
<td>83</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>24</td>
<td>48</td>
<td>3</td>
<td>17</td>
<td>21</td>
<td>66</td>
</tr>
<tr>
<td><strong>Reporting findings (initial agreement index)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are quotations (of at least one sentence in length) from participants included in the research report? (( \kappa = .62 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No quotes</td>
<td>12</td>
<td>24</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1 = 1 to 4 quotes</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>2 = 5 to 9 quotes</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>17</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3 = 10+ quotes</td>
<td>30</td>
<td>60</td>
<td>13</td>
<td>72</td>
<td>17</td>
<td>53</td>
</tr>
<tr>
<td>Do the authors include extended quotations (five or more lines of text)? (( \kappa = .71 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = None</td>
<td>16</td>
<td>32</td>
<td>3</td>
<td>17</td>
<td>13</td>
<td>41</td>
</tr>
<tr>
<td>1 = 1 or more</td>
<td>34</td>
<td>68</td>
<td>15</td>
<td>83</td>
<td>19</td>
<td>59</td>
</tr>
<tr>
<td>Are identifiers included with quotations? (( \kappa = .75 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = No</td>
<td>28</td>
<td>56</td>
<td>8</td>
<td>44</td>
<td>20</td>
<td>63</td>
</tr>
<tr>
<td>1 = Yes</td>
<td>22</td>
<td>44</td>
<td>10</td>
<td>56</td>
<td>12</td>
<td>38</td>
</tr>
</tbody>
</table>

**Note.** All dimensions are nominal codes, except for time spent with participants, which is continuous; chi-square tests are for significant differences in category frequency for early and late studies (early indicates studies published in 1989 through 1996; late indicates studies published in 1997 through 2003). Agreement indices are Cohen’s kappa for 11 nominal scales and the intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979; ICC[3, 2]) for the continuous time spent dimension.
characteristics in Table 1. The two groups of studies did not differ systematically on any of these variables (all ps > .10). The median time spent with participants was somewhat larger in mixed studies ($M_{dn} = 90$ min) than the time spent in pure QUAL studies ($M_{dn} = 54$ min). Because all mixed studies contained substantial QUAL components, and because they did not differ systematically as a group from the pure QUAL studies, we conducted all analyses on the full set of 50 studies.

**Conformity of Published Studies to QUAL Ideals, 1989–2003**

**Research focus.** The coders who analyzed the data and formulated the results had rather limited acquaintance with research participants in this set of studies. In fewer than 20% of the studies did a majority of coders have contact (either interaction or direct observation) with at least one study participant. In 28%, no coder had direct contact with participants. In the vast majority (88%) of studies, the primary data available to coders consisted of transcripts of interview conversations (or, in some cases, written responses by participants to some form of open-ended questionnaire). Thus, for most coders in most studies, there was no access to visual or paralinguistic cues that might enhance understanding of the meaning of these responses to the participants.

Another (crude) measure of acquaintance is the average amount of time spent with each participant by those study personnel who do have direct contact. For the 41 studies that specified amount of time spent, the median was 60 min per participant. Three of the 50 studies (6%) incorporated sustained observation of participants on multiple occasions, with a cumulative time spent of more than 3 hr (180 min; much more, in 2 of these studies). Two such studies (Hill et al., 2000; Morrow & Smith, 1995) included observations of multiple focus groups, as well as interviews with participants outside the therapy context. One (Wampold et al., 1995) involved interviews and direct observation of participants in laboratory groups over an extended period. But the modal study involved a single interview (or a single interview and a brief follow-up) as a basis for understanding of the experiential worlds of the participant.

**Research setting.** In contrast to the ideal of encountering participants in their natural environment, 80% of these QUAL studies involved either no contact between participants and researchers or contact in an artificial environment (such as a phone interview or an interview in a location convenient for the researcher). Thus, the data for these studies generally included participants’ verbal descriptions of their experiences and behaviors but little opportunity for researchers to observe how the participants actually behaved in a naturalistic setting. Exceptions to this trend included the three studies just mentioned, which incorporated extended observation or interaction with participants across multiple settings. Other exceptions included studies of the psychotherapy process, in which coders had access to audiotapes or (typically) transcripts of interactions in this naturalistic environment, sometimes supplemented by postsession reflections from participants in an interview setting. For example, Rennie (1994) analyzed data from transcripts of actual therapy sessions as well as postsession interviews.

**Researcher role.** An important trend in QUAL research published in *JCP* is the use of auditors as an adjunct to the main coder group. The goal of this practice (Hill, Thompson, & Williams, 1997) is to enhance confidence in the reproducibility of findings (e.g., identified themes, characterization of these themes as more or less typical in the sample) by showing that they are replicated in the main (usually with some necessity to resolve disagreements) by an independent group of coders. Although at least some of the primary coders usually had contact with one or more research participants, the auditors typically operated at a further remove, almost always relying on transcripts or other written materials as their sole source of data. More than two thirds (68%) of the studies we surveyed used auditors as a check on the coding process and incorporated the audit team’s feedback into their published results.

In a majority (58%) of published QUAL research, the authors included at least a brief reflection on their experience with the subject under study and any preconceptions about what the findings were likely to be. This practice is recommended for authors of QUAL work as a means of conveying the lens through which the research team approached the data (Morrow & Smith, 2000). In addition, in nearly half (48%) of these studies, researchers explicitly mentioned the desirability of minimizing or eliminating researcher biases, a goal that may be more congruent with QUAN than QUAL research cultures. (Note too that an even larger proportion of these studies included auditors, indicating at least an implicit endorsement of this objective.)

**Reporting findings.** Most research reports based on this set of studies made an effort to include direct quotations from participants, to convey the language and thought processes participants used in making sense of their experience. It is interesting to note that, even in this set of qualitative and mixed-method studies, nearly one quarter (24%) included no quotations as long as a full sentence (although virtually all quoted words or phrases to give some sense of participants’ word choices). Most studies (68%) included extended quotes (which we defined arbitrarily as taking up five or more lines of printed text), which give a more unadulterated picture of participants’ thinking than shorter excerpts selected by the researcher. And a sizable minority (44%) of studies included identifiers with quotations (and sometimes one or more extended case studies), which also allow interested readers to piece together a more complete sense of the experiential worlds of participants.

**Trends Over Time**

In ancillary analyses, also presented in Table 1, we compared studies from the early (1989–1996) and late (1997–2003) halves of the time interval under investigation. Our aim here was to examine temporal trends in QUAL studies published in *JCP*. It could be that early, pioneering QUAL investigations were more likely to conform to traditional QUAN validity criteria as a way of getting a foot in the door in a discipline in which research has been traditionally viewed from a QUAN perspective. Presuming that QUAL studies and methods gained greater acceptance over time, we hypothesized that more recent studies would more closely approximate QUAL ideals than did the earlier, ice-breaking publications.

Type II error rates are likely to be high for these analyses, even for time trends that may be of practical importance, because of the relatively small number of available studies. To reduce the number of Type II errors, we interpret probability values less than .05 as marginally significant. This terminology puts readers on notice that the “marginal” trends should be interpreted with caution but also that rigid adherence to the customary alpha
level of .05 may result in overlooking suggestions of important trends in publication practices in this literature.

Research focus. Evidence for this hypothesized trend is mixed in the first set of categories. Coders in recently published studies (81%) were marginally more likely than those in early investigations (56%) to have direct contact with at least one of the research participants, $\chi^2(1, N = 50) = 3.77, p < .10$. This finding clearly points to closer adherence to QUAL principles among later investigations. However, among those studies in which coders did interact with participants, phone interviews were nearly as common as face-to-face interactions in recent publications (38% and 44%, respectively), unlike in early studies (11% and 44%, respectively), a difference that is also marginally significant, $\chi^2(2, N = 50) = 5.58, p < .10$. Because phone conversations are a more remote form of interaction than face-to-face conversations, this trend appears to be in the opposite of the hypothesized direction.

On the variable of time spent with participants, the low value ($Mdn = 0 \text{ min}$) for early studies is attributable to the large proportion of these studies in which coders had no contact with participants, as just discussed. When “no contact” studies are excluded, median times for early and late studies are relatively comparable—approximately 100 min and 70 min, respectively.

Research setting. Recent studies (12%) were marginally less likely than early studies (33%) to observe or interact with participants in a naturalistic setting, $\chi^2(1, N = 50) = 3.13, p < .10$. This represents a step away from the principle that QUAL research should be conducted in the field rather than the laboratory.

Researcher role. Published reflections on authors’ biases and preconceptions about the research topic (a practice recommended for QUAL researchers) were somewhat more common in late studies (66%) than early ones (44%), although this difference was not statistically significant. Authors of recent studies were significantly more likely than those of earlier studies to make use of independent auditors to authenticate study findings (81% to 44%), $\chi^2(1, N = 50) = 7.17, p < .01$. Recent authors were also more likely to assert the desirability of minimizing or eliminating the effects of author or coder biases on published findings (66% to 17%), $\chi^2(1, N = 50) = 11.06, p < .01$. As noted above, both the goal of eliminating bias and the use of auditors in the service of this goal are arguably more consonant with QUAN than QUAL research standards. Thus, it is significant that these practices are on the rise in recent published studies. We return to this issue in the Discussion.

Reporting findings. Early studies were marginally more likely to include numerous quotes than late studies (72% vs. 53%); more important, only 1 of 18 early studies (6%) was marked by a complete absence of sentence-length quotations, compared with 11 of 32 recent studies (34%), $\chi^2(3, N = 50) = 7.54, p < .10$. Similarly, early studies were marginally more likely to include extended quotations (83% to 59%), $\chi^2(1, N = 50) = 3.04, p < .10$. Thus, the trend over time appears to be away from the QUAL ideal of rich descriptions that facilitate readers’ access to the experiential world of research participants.

Discussion

In this study, we examined 15 years of QUAL research published in JCP to ascertain the fidelity of these studies to QUAL principles on four dimensions on which QUAL and QUAN research practices are expected to differ (research focus, research setting, researcher role, and reporting of findings). Generally speaking, the fidelity was not high. Data analysis in the modal study was conducted by a group of coders, most of whom had no direct interaction with (or even direct observation of) any of the research participants. Instead, they worked from written records of participants’ verbal reports—transcripts of a 60-min interview. In contrast to the QUAL ideal of extensive interaction or observation in a naturalistic setting, this interview was typically conducted in a setting convenient for the researcher or, when this was not practical, by telephone. Typically, one or more coders also served as interviewers, so that they could bring information not available in the written record (e.g., recollections of visual or paralinguistic cues) to the attention of the primary coder group, who could, in theory, incorporate this additional, vicarious information into their interpretations of the transcripts. Typically, however, the findings of the primary coder group were audited by a second (smaller) group of coders, with the ostensible aim of removing collective biases and enhancing replicability of findings. These auditors had no way of accessing information beyond the written transcripts.

The strongest fidelity for these 50 studies was to the QUAL standards for presentation of research findings. Typically, researchers included direct quotations from the participants, and a sizable minority included identifiers with each quotation that gave readers a richer portrait of study participants.

We also examined trends over time on the assumption that pressure to abandon native QUAL cultural practices in favor of entrenched QUAN practices may diminish as QUAL methodologies become more familiar and (by inference) more accepted. In fact, on most of the dimensions on which recent studies differed significantly (or marginally significantly) from early studies, the trend was one of decreasing fidelity over time (i.e., early studies showed greater conformance to QUAL ideals than later ones). Thus, the overall trend over time can be described as one of movement away from core QUAL principles.

Implications

Advocates of methodological diversity in psychology (e.g., Denzin & Lincoln, 2000; Elliott et al., 1999; Havercamp et al., 2005; Hoshmand, 1989) pay careful attention to philosophy of science in critiquing positivist and (to a lesser extent) postpositivist approaches and in advancing the case for QUAL inquiry. They concur that QUAL studies are grounded in different epistemological assumptions than QUAN studies, and that the field will benefit by encouraging scholarship from both idiographic and nomothetic perspectives. These different ways of knowing provide multiple perspectives for understanding human experience, and both can inform psychological theory and practice.

A question raised by the present findings is whether the chances to realize this potential may be diminished if QUAL investigations conform more closely to the dominant methodological paradigm rather than being conducted and reported from an authentically qualitative (or idiographic, as we have described it in this article) perspective. Our findings suggest that, although QUAL studies published in JCP embody some of the ideals of traditional idiographic inquiries (e.g., small sample sizes, intensive analysis of text), in many other respects they conform to principles more characteristic of traditional QUAN research (e.g., concerns about
minimizing researcher subjectivity and documenting reliability or reproducibility of findings). It might be argued that these studies incorporate the best of both QUAL and QUAN traditions and, hence, constitute a beneficial synthesis of traditional QUAN and newer QUAL approaches. A different, and more worrisome, interpretation is that these deviations from core QUAL principles represent a sort of compromise with the reigning (QUAN) norms that do not fare well according to the standards of either tradition and, thus, may diminish the credibility and impact of published QUAL research. We consider two possible negative consequences of the paradigmatic admixture documented here: overclaiming and underachieving.

Overclaiming. When QUAL researchers relax their adherence to alternative research paradigms (e.g., constructivist, critical theory; Ponterotto, 2005a), and judge the merits of their findings instead according to traditional QUAN criteria, they may begin to make inappropriate claims for their findings. For example, it was not uncommon for QUAL articles in our sample to include a Discussion section with a title such as “Implications for Practice,” suggesting that findings should generalize to clients sharing some characteristics with study participants. Although traditional in QUAN articles, such considerations are out of place in purely QUAL research, in which generalizability and objective truths are neither sought nor expected. When researchers talk as if QUAL findings constitute scientific evidence for generalizable psychological theories, they distort the paradigmatic framework that grounds such inquiries, and they run the risk of misrepresenting findings for readers who may not be informed about legitimate standards for scientific inference.

Examples of this tendency to heap onto QUAL findings an inferential load that they are ill-equipped to bear are beginning to emerge in the literature on evidence-based practice. For example, Chwalisz (2003) has argued for “an expanded view of evidence” in which “all types of evidence [including both QUAN and QUAL studies] make equally valid contributions to understanding” (p. 502). Other authors have urged a similar expansion of the criteria by which psychological treatments are judged to have been empirically validated, with qualitative as well as quantitative investigations considered as evidence of effectiveness (Bernal & Scharrow-del-Rio, 2001; Sue, 2003), without carefully delineating the strengths and weaknesses of each type of evidence as a basis for practice. Similarly, the newly revised reviewer guidelines for JCP distinguish QUAL studies that are “discovery oriented” from those that are “confirmatory/verification oriented” (Mallinckrodt, 2006, p. 7), but they do not clarify how QUAL investigations can be used to test (confirm) nomothetic hypotheses. Indeed, the guidelines go on to specify that authors of QUAL studies should avoid implying that findings are generalizable [p. 10], which begs the question of what role they can play in verification-oriented inquiry.

Overclaiming, in the sense of claiming more for QUAL studies than they can legitimately deliver, is problematic both scientifically and ethically. Clearly, it is poor science to treat idiosyncratic findings from a QUAL study as though they confirmed theorized nomothetic relations among variables (e.g., “Treatment X has been shown to be effective for population Y”). As well, to make policy decisions or to promote interventions as “empirically validated” only on the basis of evidence of QUAL investigations would be ethically questionable. It seems imperative for advocates of QUAL methods, as well as editors and reviewers who evaluate QUAL studies for publication, to clarify the role of idiographic inquiry in furthering our understanding of people in general and psychological practice in particular. These deliberations should take into consideration the limitations as well as the strengths of QUAL approaches.

In a thoughtful discussion of these issues, Hill (2006) referred to Rychlak’s (1968) distinction between procedural evidence (which is evaluated on the strength of its coherence and consistency with common-sense knowledge) and validating evidence (which, in Rychlak’s usage, is comparable with the traditional use of the term evidence in scientific contexts). QUAL inquiries can provide procedural evidence for treatment effectiveness, whereas validating evidence should be based on QUAN studies. It would be useful to consider in more detail the roles and merits of these two types of evidence and the types of inferences and actions that are warranted on the basis of each. Obscuring the distinction between procedural and validating evidence can be costly in terms of credibility both of the researcher and of the intervention or theory under study—and ultimately, if such claims become prevalent, of counseling psychology as a discipline.

Qualitative postpositivists? Connected with this theme of overclaiming is the recent trend on the part of some QUAL researchers to claim that they are conducting their inquiries in accordance with a postpositivist philosophy of science. In the 2005 JCP special issue on qualitative methods, Ponterotto (2005a) acknowledged this trend (albeit with some lack of enthusiasm), and Morrow (2005) afforded it legitimacy by proposing separate criteria for evaluating the trustworthiness of “postpositivist” QUAL investigations. From the viewpoint of a postpositivist, it is difficult to make sense of the claim that QUAL investigations can be conducted within this framework. QUAL studies, with their small sample sizes, unstandardized measurement procedures, and lack of experimental manipulation or statistical control, are simply not equipped to make the rigorous examination and elimination of alternative explanations that are central to the postpositivist’s understanding of the scientific method (cf. Shadish et al., 2002). One is left with the impression that the claimants would like to have their cake and eat it too—that is, they would like to be able to draw the type of confirmatory conclusions (based on tests of a nomothetic hypothesis) that characterize interpretations of findings in postpositivist research without the necessity of attending to the research design and measurement considerations that ground such claims.

Not every commentator celebrates this trend toward imposing QUAN standards on QUAL investigations. Ponterotto (2005a) lamented what he described as the “postpositivizing” of QUAL work. Merrick (1999) also mounted a spirited defense of QUAL studies against the need to justify claims of reliability and generalizability of findings (see also Morrow, 2005). Stiles (1997) questioned seeming inconsistencies in the application of postpositivist validity criteria and associated methods to QUAL research and skillfully articulated the potential cost of such measures, which tend to undermine processes that are central to the goals of QUAL work. If methodological trendsetters encourage researchers to straddle QUAN and QUAL norms, and editors and reviewers endorse this movement, there is every likelihood that the resulting products will not be strong by the standards of any paradigm.
**Underachieving.** If QUAL researchers in psychology come to believe that they can (or must) uphold their findings as unbiased descriptions of participant experiences, provide evidence of the reliability of their major conclusions by replicating them using judges who have never met these participants, and defend the adequacy of their interpretations as tests of hypotheses about relations between variables in a broader population, it seems likely that these premises will have a deleterious effect on the framing, analysis, and reporting of QUAL research. Just as they inevitably fall short relative to the standards QUAN researchers use to justify such claims of validity, reliability, and generalizability, such studies will also fare poorly relative to the standards of their native QUAL culture. If we teach graduate students (in the classroom and by example) that QUAL inquiry consists predominantly of intensive analysis of transcripts of 60-min interviews, will the resulting studies fulfill their promise as in-depth explorations of the lived experience of participants, coconstructed with researcher-observers? If the emphasis in QUAL training and manuscript review is on documenting reproducibility of a set of coded categories, will reporting of these categories tend to take precedence over rich descriptions of individual experience and thoughtful reflections on the meanings of these experiences?

If it becomes the consensus of the QUAL research community in counseling psychology that conformity to postpositivist standards is achievable, acceptable, and even expected, we believe that the potential of QUAL methods to inform research and practice will be drastically curtailed. The necessity of conforming to traditional QUAN standards may diminish (a) the QUAL researcher’s focus on intensive interaction with and observation of participants in their natural contexts; (b) the flexibility to explore tangential material or to probe for greater depth of insight in ways that may deviate from the interview protocol; and (c) the permission to portray the intersubjective journey between researcher and participant with insight and artistry, rather than producing sanitized consensual codes. In short, it may lead to an underrepresentation in published research of the very strengths that QUAL methodologists (and we) believe can contribute to a broader perspective on the human experience in the research literature in counseling psychology.

**Concluding Thoughts**

We suggest that the publication record of QUAL researchers in *JCP* demonstrates a tendency in counseling psychology to assimilate to the dominant QUAN norms, altering customary QUAL assumptions and goals to achieve greater congruence with those of the reigning postpositivist paradigm. This tendency appears to be increasing over time, so that there is a danger that counseling psychology is creating a type of QUAN–QUAL hybrid that may not really accomplish the goals of either approach. There is evidence of tension around these issues within the qualitative research community, where some advocate methods that “combine the best elements from both [QUAL and QUAN] traditions” (Hill, Williams, & Thompson, 1997, p. 613), but others evince caution regarding the philosophical consistency and epistemological viability of such hybrid methods (e.g., Stiles, 1997).

An empirical investigation can tell us something about how things are but not how they should be. Decisions about how QUAL methods may best contribute to knowledge in counseling psychology must be grounded in epistemology—that is, beliefs about what sorts of knowledge are useful and for what purposes. Our findings raise the question whether the potential of QUAL methods is being fully realized in counseling psychology and suggest the need to clarify the potential role of QUAL inquiry in informing science and practice in counseling psychology. On the basis of our review of current publication practices, attention also seems warranted to the nature of evaluative criteria and editorial procedures that will encourage QUAL researchers to do the type of work most likely to make this potential a reality.

**References**


Received April 6, 2006
Revision received January 3, 2007
Accepted January 16, 2007