

Helping Child Welfare Workers Learn Interviewing Skills

Colleen Friend

Case Western Reserve University

Cleveland, Ohio

Statement of the Research Problem

If the children are our future, then the state has an interest in the creation of a productive citizenry from the ranks of its children. Thus, it has a motive for protecting its youngest citizens from those who would do them harm. The latest national incidence study of child abuse and neglect identified over 1.5 million victims of child abuse or neglect and indicated that 78% of the perpetrators were parents, 10% were other relatives, and the remainder were unrelated (Sedlack & Broadhurst, 1996). With the passage of the Child Abuse Prevention and Treatment Act (CAPTA) of 1974, the federal government took the initiative to establish a model statute for state child protection programs that mandated standard methods for reporting and investigating child abuse and neglect (Costin, Karger & Stoesz, 1996).

Some scholars have argued that the vagueness of both the definition of child abuse and the reasonable suspicion reporting threshold cause dramatic over-reporting (Besharov, 1987); in fact, national incidence surveys uncover that all the time nearly 50% of child maltreatment victims are unknown to Child Protective Services (CPS) agencies (Zellman & Faller, 1996). (The National Incidence Study used community professionals as “lookouts” for maltreated children.) Their discovery that there were more maltreated children than those reported to CPS, lends support to Finkelhor’s (1993) contention that the essential problem is still underreporting, not overreporting.

Central to the federal mandate to investigate child abuse and neglect is the child welfare worker’s ability to interview the parents who come to their attention. But there is an inherent adversarial stance between the interviewer and interviewee, since these interviews are the mechanism the state uses to gather information in the exercise of its social control function. Parents who come to the child welfare system’s attention correctly perceive the stakes as high, despite the fact that few children are removed from their homes in these encounters. Practice wisdom validates the likelihood that the child welfare worker will be given very limited disclosures about facts or feelings from defensive parents, making these interviews difficult to conduct. In this encounter, the interviewer must explain the purpose, build rapport, ask a series of difficult questions,

deescalate anger, and manage his/her own emotions. The challenge of balancing all these tasks simultaneously often hijacks the ultimate goal which is to determine if the child is safe or in need of protection. The temperament of the interviewer must be accounted for as well. An interviewer who avoids conflict or becomes overly aligned with parents could run the risk of making a false negative assessment on safety issues, while an interviewer who becomes emotionally engaged with hostile parents could conversely make a false positive assessment of the same issue. Some child welfare workers seem to have a talent for minimizing the power differences, others are unaware, unwilling, or poorly skilled at doing so. It appears that this talent is a skill set that has not been well identified, trained, practiced, or even evaluated. The only published research on child welfare worker interview skills has focused on interviewing child sexual abuse victims with adults acting as though they were children (Brittain, 2000; Freeman & Morris, 1999; Stevenson, Leung & Cheung, 1992). Virtually no studies have been published that measure what child welfare workers actually do with parent-clients. Public child welfare (PCW) was once the exclusive domain of Masters of Social Work (MSW) trained social workers (National Association of Social Workers, 2000). Over time, however, the educational and experience requirements for child welfare workers have been considerably reduced, with no MSW requirement at this time, and with the job classification transformed into a generic title that lacks professional specificity or identity.

Given that the public child welfare workforce typically has little experience, and a variety of educational backgrounds, it is not clear that reinstating the MSW requirement would improve the quality of interviewing in the child welfare system. Despite the tradition of training to practice specific skill sets in MSW programs, social work teachers, researchers, field supervisors, and clients have lamented the little attention given to the practice and evaluation of interviewing skills (Badger & MacNeil, 2002; Carillo, Gallart & Thyer, 1993; Schinke, Smith, Gilchrist & Wong, 1978; Linsk & Tunney, 1997). MSW students themselves have reported feeling ill-prepared to negotiate the complexities of the interview situation (Carillo, Gallart & Thyer, 1993; Schinke, Blythe, Gilchrist & Smith, 1980). Although there has been a promising coordinated effort to draw down one of the last federal entitlements (Title IV-E dollars) for the specific preparation of public child welfare workers in MSW programs, the jury is still out on the overall effectiveness of this approach to reprofessionalizing the workforce (California Social Work Education Center, 1999).

It seems that the most effective remedy to address the problem of the deprofessionalization of the child welfare workforce is on-the-job training, yet such an approach brings its own set of complexities. With the availability of Title IV-E funding to support training, there have been a variety of trainings offered to the workforce; however, their content, including the effectiveness of the practice or transfer of knowledge, has been relatively unevaluated (McDonald & McCartney, 1999). Although it is suspected that some of this information is available within state agencies and among privately hired trainers, in general it is unpublished and intentionally confidential. In 1998, at least 25 state child welfare agencies were operating under consent decrees specifically for poor service delivery (Schwartz & Fishman, 1999), giving rise to one plausible explanation for the information gap. In addition, the work environment presents a heavy caseload demand, which is not conducive to the practice of skills

learned in training. Supervisors do not have time to observe, reinforce and retrain their subordinates (Freeman & Morris, 1999). Trainers are challenged to address a wide variety of academic backgrounds and fledgling skills in a workshop format and timeframe.

Having escaped public scrutiny and pressure to conduct research on its practices for so long (Gelles, 2000; Lindsey, 2003), the public child welfare system has now been put on notice by the National Academy to show itself to be more accountable by instituting outcome-oriented, consumer-sensitive and research-based methods (Chalk & King, 1998). The Adoptions and Safe Families Act (1997) (ASFA) has made states responsible and competitive in meeting certain outcomes.

Against this backdrop of intermittent public interest, power disparities, deprofessionalization, and a new outcome/customer orientation stance, this study was completed. This study examined, for the first time, exactly what public child welfare workers actually do in the course of an interview with a parent. The Los Angeles County Department of Children and Family Services (DCFS) was responsive to a request to conduct the proposed research, with a small sample of its workforce. A voluntary sample of workers was offered a one-day training using several approaches to learning interviewing skills discussed in detail below. Standardized clients (SCs) were used to help evaluate the workers' transfer of training skills. Standardized clients (actors/actresses) portray a client and allow for practice and evaluation without the concomitant risk of harm to real clients. An instrument was developed to capture and measure the interaction between child welfare worker and the SC. SCs were asked to provide feedback regarding the worker's effectiveness. Overall, it was hypothesized that training will make a significant difference in skill demonstration. Beyond the hoped for difference, levels of clinical change was assessed. The study's findings may become part of the public child welfare agency's efforts to improve and extend its training, retain workers, and address consumer complaints.

Research Background and Hypotheses

Five key propositions from the National Research Council's (NRC) volume, *How People Learn*, supported this research and were then linked to three theories that underscored this experiment. Baldwin and Ford (1988) noted that their model of trainee characteristics, training design, and the work environment were critical "inputs;" it captured the emerging state of training/transfer research almost twenty years ago. Subsequent reviews of training and transfer literature supported the ascendancy of Bandura's Social Cognitive Learning Theory (SCLT), including self-efficacy, as demonstrating usefulness in facilitating skill retention.

Bandura (1977; 2001) defined self-efficacy as the belief in one's own capacity to organize and execute the courses of action required to manage prospective situations. He proposed that an individual's *expectations* about behavioral reinforcements influence behavior more than actual previous reinforcement. This revolutionary concept emphasized beliefs and perceptions and challenged reliance on strict behaviorism. Integral to this departure was Bandura's (1977) emphasis on personal evaluation as a

means of positive reinforcement. He hypothesized that self-respect, self-satisfaction, and belief in one's own competence are all goals and motivations. In essence, it is self-efficacy that links an individual to performance.

James Zull (2002) takes David Kolb's Experimental Learning Theory (1984; Kolb, Boyatzis & Charalampos, 2000) a step further, matching regions in the brain and their functions to Kolb's learning cycle. According to Zull (2002), the key to all of this is the front and back transmission of brain activity from Kolb's learning cycle that mimics the brain's cycle. When utilized in a balanced approach, learners convert data into their own ideas and actions, experiencing this conversion as learning (National Research Council, 2002; Zull, 2002). Zull maintained that in learning, transfer is about taking the physical action step of testing. Until we do that, all we have acquired is merely fanciful conjecture; action makes the learning cycle complete (Zull, 2002). Zull also notes that testing helps the learner fill in the details of how to navigate between learning gaps. Transfer takes time for contemplation, action, and even random reaction. According to Zull, learners can have a brain based emotional reaction to the teacher that impacts motivation; thus the teacher must strike an alliance that does not engender fear. The teacher must challenge them to think in the classroom and in other novel situations for transfer to have a future.

The training was delivered utilizing SCLT techniques and Zull's approach to learning. The Baldwin and Ford model of transfer and Zull's Biology of Learning helped identify factors before, during, and after the training that may have influenced the training's transfer. Because a new standardized measure was developed for this experiment, classical measurement and test theories guided the establishment of validity and reliability for the instrument (DeVellis, 1991).

The specific questions asked were:

- Does a brief interview training for PCW workers, using SCLT, lead to skill transfer in a demonstration with a standardized client?
- How can trainees' interview skills in a demonstration with a standardized client be measured?
- If skills are transferred, is there a particular pattern in how that takes place that might be based on trainee characteristics, training design, or the work environment?

The hypothesis, related to each question, was:

- Interview training, using SCLT, will significantly improve interview skill performance in a sample of PCW workers.
- An instrument designed to measure interview skill demonstration will demonstrate validity and reliability within a small developmental sample.
- The research will reveal patterns of skill demonstration between highest and lowest level skill demonstrators that will inform future training efforts in the areas of trainee characteristics, training design, and work environment.

The last hypothesis was examined both quantitatively and qualitatively. A conceptual content analysis was used to determine what skills set an expert apart from a novice. This kind of content analysis relies on theory developed by Krippendorff (1980) and was discussed in the Colorado State University's (2003) web-based publication on content analysis history and methodology. These relationships are outlined in Table 1.

Methodology

This experiment (also referred to as the core experiment) used a quasi-experimental model, with a pretest, intervention posttest ($O_1 X O_2$) design. A small volunteer sample of Public Child Welfare workers was recruited from one region of a large public child welfare agency. The independent variable was the training and the dependent variable was the demonstrated interview skill level. The research utilized two SCs who had been previously employed in the UCLA Medical School's Identified Patient Program. Thus they readily adapted to several hours of training in order to reliably represent the same allegedly battered mother who was "reported" to the local public child welfare agency. The SCs were matched on gender, ethnicity, age, and the ability to stay in character. They were Caucasian females in their early thirties. Two vignettes were utilized, rated at the moderate level by three researchers. Both the vignettes and the SCs were switched at the posttest, to assure trainees (also referred to as subjects) would not become overly familiar. Two teams operated simultaneously, resulting in data collection on all subjects in a concise timeframe: the same process was repeated at time two, less than two weeks after the training intervention. Subjects were rated by two raters in the interaction with the SC. The raters were the author and three Title IV-E stipended MSW students, who trained together for 12 hours, in order to ensure acceptable reliability.

Overall, five instruments were used to collect data. They are summarized in Table 2 and briefly described here. First, demographic data on each subject was gathered, on the Demographic Data form (DD). Next, the Phase II, Part I Questionnaire (PPQ) evaluated the trainees ability to assess what was happening, what their observations were based on, and what specific things they planned to do. The actual interaction of the subject and the SC was coded on The Child Welfare Domestic Violence Interview Skills Scales (CWDVISS), an instrument developed for this study. It was constructed in accordance with DeVellis' (1991) seven step process. Its reliability and validity was piloted with a separate and small "developmental sample" of actual workers. Ultimately, this instrument included 14 fields, or "skill clusters," that specifically pertain to skills needed for interviewing with cases involving the intersection of domestic violence and child abuse. These skill clusters were scored quantitatively, with points assigned for repeated, specific skill demonstration, and points subtracted for repeated, specific errors. The CWDVISS served as an anchor for a later qualitative content analysis, which included a post hoc analysis of the audiotaped interview. The standardized client (SC) used the Patient-Physician Interaction Form (PPIF) to rate the subjects. Although there were no published reports of this instrument's validity or reliability, its use is widespread in Identified Patient Programs in California Medical Schools. This instrument has seven fields rated on a five-point Likert scale. It measures

the patient's global level of satisfaction during their interaction with the medical student. It was expected that a correlational analysis would reveal a positive relationship between the PPIF score and the CWDVISS. It was important to approach the experiment with concurrent measures to provide a mechanism for feedback to the worker and involve the SC as a client proxy. In addition, a validity check for the CWDVISS in the developmental sample was constructed by compiling three subscales from Finn and Rose's (1982) Interview Skills Role Play Tests (ISRPT), referred to as the Finn and Rose's Subscales (FRS). While the "borrowing" of the ISRPT's subscales does not retain the original instrument's established reliability and validity, what was being sought was a broad check on validity through establishing construct validity.

There was a challenge in recruiting subjects for these two samples: The first is referred to as the "developmental sample" for testing the CWDVISS's reliability and validity; the second is the sample for the core study. Subjects were recruited by email and given the following incentives: employment training credit for all levels of participation, a domestic violence book, and a feedback letter (with audiotape if requested). Subjects were assured that their names, individual scores and audiotaped interviews would be confidential. Ultimately, these were two small convenience samples, consisting of 6 child welfare workers in the developmental sample, and 15 child welfare workers in the core study. The original core study was a set of 19, but two workers dropped out at the posttest, citing casework emergencies, and two sets of records became detached, rendering them unusable. In sum, the core study sample was 60% male and 40% female, with 46% African-American workers, 34% Caucasian and 20% Latino/a workers. The higher representation of African-American males in this sample was typical of the workforce in that region (sample frame) at the time of the study.

The curriculum for this training intervention was adapted from a training previously published by the author as: *Assessment and Intervention Approach to Domestic Violence Cases Involving Children: An Innovative Training Program for Child Welfare Workers* (Friend, Mills, Hoang, Maxwell & Rubin, 1999). The curriculum was also summarized in a subsequent publication in the *Social Workers' Desk Reference* (Friend & Mills, 2002). Both publications contain a shorthand version of the original assessment instrument; its content was the foundation of the 14 skill clusters contained on the CWDVISS. Table 3 outlines process and content points that enhanced the original training's content to become this experiment's intervention.

What follows is the data collection procedure for both components (core and developmental sample) of the study. For the developmental sample, data were collected in an identical manner as described in the core study below with these exceptions: subjects were coded by two raters in only one observation, without the training intervention, and the audiotape of the interaction was coded that same evening by the same raters with the FRS as an additional validity check on the newly developed CWDVISS.

In the core study, once subjects signed consents, they were given the child abuse hotline/referral on the client. They then viewed a videotape of the SC reacting to the DCFS hotline report. They were asked to fill out the PPQ, which asked three preliminary questions to help subjects formulate a cognitive "to do" list. They were then told they would have up to one-half hour to interview the identified patient, be observed by two

coders, and audiotaped. Subjects were told the audiotape would be made to allow for supplemental coding after the interview concluded. Two teams of two raters each operated simultaneously and separately, rating the trainee's interviewing skills on the CWDVISS. Groups were identified as "A" and "B" so that the standardized client (SC) could be switched at time two. The literature review advised that previous experimenters detected a "familiarity effect" when subjects encountered the same SC at time two. As a protection against such threats to internal validity, this switch was made.

In order to determine if there were unique or shared patterns of skill acquisition, the highest scoring (both pretest and posttest) subject on the CWDVISS and the most improved (from pretest to posttest) subject were examined quantitatively and qualitatively. The assumption was that the former represented the best preexisting (i.e., expert) skill and, the latter, the most improved (i.e., novice) skill. Here the pattern of their CWDVISS scores, their scores on the PPQ and a content analysis of the audiotape were conducted. The PPQ "to do" list was rated in accord with Cournoyer's practice rubric (Cournoyer, 2004). Next, two researchers coded three randomly chosen tapes; two were used to practice coding, the third was a test of reliability. The number of actual agreements divided the number of possible agreements. For this content analysis, the reliability was .87, which is consistent with Krippendorff's (1980) range of acceptable reliability.

Results

Because this was complex research utilizing a developmental sample ($n = 6$) for instrument reliability and validity, and another sample ($n = 15$) as the core study, the findings will be recapped here in an effort to achieve clarity. First, a pilot test of reliability and validity of the CWDVISS with a small developmental sample found that the instrument's reliability, calculated between the two raters, was .85 with a corresponding mean alpha of .88. Construct validity of the CWDVISS was established by determining the extent to which correlations among all three instruments (CWDVISS, PPIF, and FRS) led the researcher to believe that the CWDVISS scores behaved the way they were expected in relationship to established measures of other constructs. These bivariate correlations are in the moderate range and statistically significant at $p < .06$. These findings are summarized on Tables 4 and 5. In the core study, t -tests conducted on participants' CWDVISS pretest and posttest scores demonstrated that the approximate 28-point difference was statistically significant at $p = .01$. This suggests that the independent variable (training) had a positive impact on the subject's interviewing skills. Effect size calculations, using Cohen's d , were estimated at 1.05. These results are summarized on Table 6. This reflects a high degree of impact or effectiveness attributable to the independent variable, which was training. The three subjects who had previously taken domestic violence training did significantly better at the posttest than their counterparts without this training. Caution should be exercised when interpreting this finding because of the small sample size. None of the other differences in mean scores identified in Table 6 revealed a statistically significant result when the data were run by gender or ethnicity. It is important to note that all 7 of the subjects with less than three years experience received the lowest scores. All of his or her scores improved at the posttest; this was not true of every experienced subject in the sample.

In addition to the quantitative analysis just described, two subjects whose scores were extreme, i.e., lowest at baseline to most improved (subject 2) and highest baseline (subject 9) were probed both quantitatively and qualitatively, revealing different patterns of skill acquisition in the posttest observation. In sum, this comparison can be summarized in this way: engagement (Field 1) and listening (Field 9) are high scores for both subjects in both observations, and conducting safety planning is a low score for both. Subject #2 scored well on giving options at the posttest, altogether suggesting that he might be using the mnemonic (cognitive) strategy made up by the trainee group, i.e., LEGO (Listen, Explain, Give Options). Both subjects' low score on field 12 (safety planning) may reflect that this skill cannot be acquired with the intervention the study offered, or it may be consistent with the difficulty many helping professionals have demonstrating this skill (Davis, personal communication, 2002). Subject 2's dramatic improvement in the posttest reveals a shift of increased skill demonstration in establishing a partnership and explaining/giving options. It might mean that subject 2 began to share power and draw the client into participating, as this is what these two skills have in common. On the other hand, subject 9 seems to be doing something very different. First, he focuses almost exclusively on engagement and listening in the first segment, then he waits until the second ten minute segment of the interview to shift into an escalated discussion of the presence of domestic violence and he then waits to conduct a threat assessment close to the end of the interview. His ability to develop the relationship and pace his intervention won him the highest scores from the SCs. Subject 9 was consistent in the demonstrating of nonjudgmental feedback, something subject 2 only partially demonstrated at the posttest. Despite being the highest scorer throughout, subject 9's scores slightly dipped at the posttest's second and third segments, suggesting he was experimenting with some new skills. Lastly, subject 2's continued brevity may have cost him even more points, leaving open to speculation what he might have achieved had he kept at his new skills, or paced himself in the manner set forth by subject 9.

Turning the focus to whether or not the extreme scorers were different on their execution of what they said they would do, the answer is yes. The analysis of the audiotape shows that subject 2 had difficulty making a plan at the pretest. In execution, he was scored as having errors in the pretest for being very judgmental. That was noticeably improved at the posttest, suggesting the intervention helped him. The intervention seemed to also help him come up with a better to do list; at the posttest all his items were scored as being within good practice and/or espoused in the training. Subject 9 had none of these issues at either observation; his audiotape coding revealed perfect execution of a full score plan. An examination of his demographics showed that he had been a child welfare worker for five years, while subject 2 had been hired within the last year.

Limitations and Utility for Social Work Practice

There are several key limitations to this study. First, this quasi-experimental design precludes causal statements. Second, these small voluntary samples cannot be said to be representative of the whole population; it was intended to be a pilot study.

Third, the theoretical sample of extreme scores helps us delve into the details of skill acquisition, but it may not represent the whole sample's process (Miller & Crabtree, 1992). Given those understandings, five key theoretical principles were supported. First, the active testing offered by the in-training role play and the pretest interview with the SC probably allowed the trainees to do significantly better at the posttest as a group. Second, this research may have "hit" the appropriate timeframe (less than 2 weeks between observations) for significant skill acquisition, retention and demonstration. Third, the trainer may have struck an alliance with the part of the subjects' brains that governs the emotions, allowing them to take in the training experience, and demonstrate in an anxiety generating observation, a significantly improved score as a group. Fourth, it appeared that the most improved trainee was using the metacognitive strategies such as LEGO (Listen-Explain-Give-Options) that paralleled the strong profile that both extreme scorers showed in the quantitative analysis. Lastly, the training of both the experienced and novice subjects together did expose those inexperienced subjects to advanced strategies; this could have been partially responsible for the novice subject's improved performance at the post test. This experiment lacked the resources to control for all of the multiple variables that participate in Baldwin and Ford's model, outputs, and conditions of transfer. The design of the training and small sample preclude attributing the findings to any one theoretical application alone.

Nevertheless, this pilot study is the first study with child welfare workers that actually provides a window into what they do with adult clients and how a brief training might improve that interaction. Next, the study proposed a methodology for demonstrating and measuring trainee skill with a standardized client. As shown here and elsewhere, this method for interview training can be a powerful evaluation tool to improve social work education (Miller, 2004). The instrument developed demonstrated preliminary reliability and validity, but reuse of this instrument should be preceded with factor analysis to reduce its complexity and make the coding less labor intensive. In this study it appeared that experts and novices demonstrate skills differently, thus they may learn differently, which has the potential to inform social work teaching practice. Finally, the study established the complexity of PCW interviews, and how proficiency requires nuanced and substantive skill demonstration. Although a one day training here made a significant difference, workers appear to need even more training. Given a posttest mean score of 73 out of a possible 138, this raises the issue of the relationship of statistical significance to clinical adequacy. It is worth noting that despite the lack of reliability and validity attributable to the PPIF, the SCs' mean score rating improved only one point from pre to post test. This could be an indication that the subjects' interview skills were still in need of improvement.

The inclusion of SCs in this research as a proxy for clients has implications for the future role of current or former clients in this kind of skill training and research design. It is a harbinger for the growing movement of PCW agencies toward a long delayed consumer consciousness. The Institute for the Advancement of Social Worker Research (IASWR) has recently addressed this issue in its Workforce and Accountability report, which identified agencies where families are being engaged in the ASFA outcomes review process and are becoming more aware of their rights (IASWR, 2004). Finding opportunities for workers and clients to work in mutually designed training and research partnerships participates in a form of reciprocity identified by Freire (1993) that

can help reduce the adversarial nature and power disparity that often surfaces in these interviews. Ultimately, it is the relationship between the worker and the parent that allows any appraisal of the parent's protective capacity to take place; thus interviewing to develop a relationship must be integral to PCW's broad child safety mission.

The publication of this research will coincide with Congress' weighing of legislative proposals and foundation reports for the dramatic restructuring of Title IV-E funding. An overhaul is likely to combine and "cap" direct service funds with training funds. This, in turn, could affect both PCW training and MSW education as these entities will likely receive fewer funds over time to prepare the workforce for navigating the nuanced and substantive skills described in this research. Unfortunately, as this research indicates, trainees need more, not less training. What is recommended is that PCW workers and MSW students who are preparing for PCW work be trained in ways that allow for measurable skill demonstration, and that SCs or clients be included in the training, research and outcome review process. Then universities where MSWs are trained for PCW work and PCW agencies can join forces to enhance their case for insuring ongoing stable training funds in this important public policy area.

The author is grateful to the California Social Work Education Center for providing a curriculum development grant that funded this research.

References

- Badger, L., and MacNeil, G. (2002). Standardized clients in the classroom: A novel instructional technique for social work educators. *Research on Social Work Practice, 12*(3), 364-374.
- Baldwin, T., and Ford, J. (1988). Transfer of training: A review and directions for future research. *Personal Psychology, 41*(1), 63-105.
- Bandura, A. (1977). Self-efficacy: Toward a unified theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (2001). Social cognitive learning theory: An angentic perspective. *Annual Reviews Psychology, 52*, 1-26. Available at: <http://psych.annualreviews.org/cgi/reprint/52/1/1.pdf>.
- Besharov, D. (1987). *Child Abuse and Neglect Reporting and Investigation: Policy Guidelines for Decision Making*. Washington, D.C.: American Bar Association.
- Brittain, C. (2000). The effect of a supportive organizational environment on transfer of training in child welfare programs. Doctoral dissertation (unpublished). University of Colorado at Denver.
- California Social Work Education Center. (1999). Tenth Year Anniversary Report. Berkeley, CA: Author.
- Carillo, D., Gallart, J., and Thyer, B. (1993). Training MSW students in interviewing skills: An empirical assessment. *Arete, 18*(1), 12-19.
- Chalk, R., and King, P. (eds.) (1998). *Assessing Family Violence Interventions: Linking Programs to Research-Based Strategies*. Washington, DC: National Academy Press.
- Colorado State University. Writing @ CSW: Writing Guidelines. Overview: Content Analysis. Available at <http://writing.colostate.edu/references/research/content/index.efm>.
- Costin, L., Karger, H., and Stoesz, D. (1996). *The politics of child abuse in America*. New York: Oxford University Press.
- Courmoyer, B. (2004). *The Evidence-Based Social Work Skills Book*. Boston: Allyn & Bacon.
- DeVellis, R. (1991). *Scale development: Theory and applications*. Sage Applied Social Research Methods Series, 26. Newbury Park, CA: Sage.
- Finkelhor, D. (1993). The main problem is still underreporting, not overreporting, in R. Gelles, and D. Loske (eds.) *Current Controversies in Family Violence*. Newbury Park: Sage.
- Finn, J., and Rose, S. (1982). Development and validation of the interview skills role-play test. *Social Work Research & Abstracts, 18*, 21-27.

- Freeman, K., and Morris, T. (1999). Investigative interviewing with children: Evaluation of the effectiveness of a training program for child protective service workers. *Child Abuse & Neglect*, 23, 701-713.
- Freire, P. (1993). *Pedagogy of the oppressed*. New York: Continuum Publishing.
- Friend, C., and Mills, L., (2002). *Domestic violence and child protection services: Risk assessment*, in A. Roberts and G. Greene (eds.) *Social Workers' Desk References*. New York: Oxford University Press.
- Friend, C., Mills, L., Hoang, P., Maxwell, E., and Rubin, J. J. (1999). *Interventions in Domestic Violence and Child Maltreatment: An Innovative Training Program for Child Welfare Workers*. Washington D.C.: United States Department of Mental Health and Human Services.
- Gelles, R. (2000). How evaluation research can help reform and improve the child welfare system. *Journal of Aggression Maltreatment & Trauma*, 3(1), 7-28.
- Institute for the Advancement of Social Work Research. (2004). *Workforce and Accountability: Child and Family Services Reviews – Implications for Child Welfare Practice, and IASWR Child Welfare Workforce Initiative*. (Retrieved from www.iaswresearch.org on March 24, 2004.)
- Kolb, D. (1984). *Experiential learning: Experience as the sources of learning and development*. Eaglewood Cliffs, New Jersey: Prentice Hall.
- Kolb, D., Boyatzis, R., and Charalampos, M. (2000). Experiential theory: Previous research and new directions, in R. Sternberg and L. Zhang (eds.) *Perspective on cognitive learning and thinking styles*. NJ: Lawrence Erlbaum Publishing.
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.
- Lindsey, D. (2003). *The Welfare of Children* (2nd ed.). New York: Oxford Press.
- Linsk, N., and Tunney, K. (1997). Learning to care: Use of practice simulation to train health social workers. *Journal of Social Work Education*, 33(3), 473-489.
- McDonald, J., and McCartney, B. (1999). Effective partnership models between the state agencies, community, the university and the community service providers in changing paradigms of child welfare practice: Responding to opportunities and challenges. Monograph from symposium held June, 1999, 43-72. Washington, D.C.
- Miller, M. (2004). Implementing standardized client education in a combined BSW and MSW program. *Journal of Social Work Education*, 40(16).
- Miller, W., and Crabtree, B. (1992). Primary care research: A multimethod typology and qualitative road map, in B. Crabtree and W. Miller (eds.) *Doing qualitative research*. Sage: Newbury Park, CA.
- National Association of Social Workers Fact Sheet. (2000). Available online at <http://www.naswdc.org>. Retrieved on Feb. 3, 2003.

- National Research Council. (2002). *How people learn: Brain, mind, experience and school*. Washington, D.C.: National Academy Press.
- Schinke, S., Blythe, B., Gilchrist, L., and Smith, T. (1980). Developing intake-interviewing skills. *Social Work Research & Abstracts*, 16(4), 29-34.
- Schinke, S., Smith, T., Gilchrist, L., and Wong, S. (1978). Interviewing-skills training: An empirical evaluation. *Journal of Social Service Research*, 1(4), 391-401.
- Schwartz, I., and Fishman, G. (1999). *Kids raised by the government*. Westport: Praeger.
- Sedlack, A., and Broadhurst, D. (1996). Third National Incidence Study of Child Abuse and Neglect. Washington D.C.: U.S. Department of Health and Human Services.
- Stevenson, K. M., Leung, P., and Cheung, K. M. (1992). Competency based evaluation of interviewing skills in child sexual abuse cases. *Social Work Research & Abstracts*, 28, 11-16.
- Zull, J. (2002). *The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning*. Sterling, Virginia: Stylus Publishing.

TABLE 1: HYPOTHESES SUMMARY

Question or Hypothesis	How Measured	Statistical Test	Theory at Work
Interview training using social cognitive learning theory will significantly improve interview performance in a set of PCW workers	CWDVISS (2 raters) SC ratings on PPIF	Paired <i>t</i> -tests	Social Cognitive Learning Theory with Baldwin & Ford Model; Zull's Approach to Learning
Instrument designed to measure above will demonstrate reliability within a small sample	Interrater reliability	Correlations	Classic test and Psychometric theory
Instrument designed to measure above will demonstrate validity within a small sample	Correlation between CWDVISS, PPIF and FRS Expert rater concurrence	Correlations	Classic test and Psychometric theory
In order to uncover what might account for very different scores, the highest and lowest scoring interviews will be analyzed at critical junctures to determine what might account for score differentials.	Identification of key skill clusters: 1) Engagement 9) Listening 12) Safety Planning 13) Explaining Options Conduct content analysis of audiotape	Conceptual Content Analysis	Content Analysis

TABLE 2: EXPERIMENTAL INSTRUMENTS

Name	Acronym	Purpose
Demographic Data	DD	Asked subject's age, race, education, years of experience, level of previous training. Used in both developmental sample and core study.
Phase II, Part I Questionnaire	PPQ	Asked three preliminary questions to help subjects form a plan used in the core study only. Maximum points: 18.
Child Welfare Domestic Violence Interview Skills Scale	CWDVISS	Measured skill demonstration in both developmental sample, test for reliability and validity, and in the core study; 14 fields: Engagement; Assessing for DV; Demonstrating Priority of Safety; Addressing Potential for Child Removal; Establishing a Partnership; Providing Feedback Nonjudgmentally; Inquiring about Strengths; Inquiring about Injury; Listening; Conducting Threat Assessment; Conducting Social Support Inventory; Engaging in Safety Planning; Explaining Options; Providing Resources. Maximum points: 137.
Patient-Physician Interaction Form	PPIF	Measured the SC's reaction to and appraisal of the subject's skill demonstration; 7 fields: Listening; Gathering Information; Establishing Rapport; Exploring Perspective; Addressing Feelings; Appearing Competent, Meeting Patient Needs. Used in the developmental sample test and core study. Maximum points: 35.
Finn and Rose Subscales of Interview Skills Role-Play Test (ISRPT)	FRS	Measured three subscales (verbal following from seeking concreteness, nonjudgmental responding) of the Interview Skills Role Play test. Used as a validity check in developmental sample test only. Maximum points: 90.

TABLE 3: KEY TRAINING INTERVENTIONS SUMMARIZED

- Acknowledged the tension between domestic violence and public child welfare service providers.
- Addressed potential feeling reactions (fear, overwhelmed, helpless) and normalized them.
- Recognized that higher rates of family violence exposure exist among helping professionals. Addressed how this could be a help or a hindrance in job performance. Discussed what to do if it becomes a hindrance.
- Empathized with workload/organizational demands and their impact on trainees' decision making, addressed paradox of demand to do more work.
- Identified trainer's work history/experience.
- Elicited trainees' experiences and impressions.
- Acknowledged previous academic training was probably not addressing this.
- Utilized visuals (family violence tree, heart of intimate abuse video demonstration) to explain concepts.
- Provided skill demonstration before role-play performance.
- Collaborated on using cognitive techniques to develop group's own mnemonic devices.
- Rotated role-play roles, to facilitate experiencing more than one perspective of the dilemma.
- Provided trainees feedback on strengths demonstrated in role play first and then coaching on other options.
- Solicited trainees' anticipations of benefits to using this method, and appraisal of utility; asked for negative feedback.
- Built on previous knowledge, experiences; elicited what these were.
- Provided conceptual theories/strategies for the development of a framework in the trainee: feminist theory, person-in-environment, Motivational Interviewing Principles and Stages of Change theory.
- Utilized the structure of an instrument to summarize training and guide initial role plays.
- To some extent, the pretest interview with the SC participated in the intervention because trainees "experienced" the SC's reaction to their baseline interviewing.

TABLE 4: DEVELOPMENTAL SAMPLE SCORES: FRS, CWDVISS, AND PPIF SCORES

Subject	FRS Subscale Total	CWDVISS Mean Segment 1	CWDVISS Mean Total	PPIF
1	10.00	28.00	38	4.0
2	34.00	34.50	47.5	29.0
3	84.00	34.50	108	35.0
4	25.00	26.00	81	19.0
5	45.00	31.00	53	22.0
6	40.00	31.50	31.5	13.0

TABLE 5: INTERCORRELATIONS BETWEEN INSTRUMENT SCORES

	Segment 1	PPIF
FRS	$r = .72^*$ ($p = .055$)	$r = .82^{**}$ ($p = .027$)
Segment 1 CWDVISS	--	$r = .725^*$ ($p = .052$)
Total CWDVISS		$r = .721^*$ $p = .053$

*Statistically significant at $p < .06$

**Statistically significant at $p < .05$

TABLE 6: PRETEST AND POSTTEST STUDY SCORES WITH EFFECT SIZE

CWDVISS*				PPIF**	
OBS	ID	PRE 1	POST 1	PRE 2	POST 2
1	1	13	97	19	26
2	2	8	102	8	19
3	3	47	90	29	29
4	6	36	99	7	31
5	9	103	106	35	35
6	10	49	29	0	3
7	15	53	106	17	5
8	19	48	37	20	14
9	20	31	36	19	10
10	14	48	61	11	7
11	18	39	90	24	8
12	12	75	71	27	22
13	13	27	67	18	23
14	11	81	64	19	22
15	17	24	45	9	28
Total Mean Score		45.46	73.33	17.40	18.80
Difference		27.87		1.40	
SD		25.53	27.38	9.78	10.30

$$\text{CWDVISS Cohen's } d = \frac{73.33 - 45.46}{\text{pooled } SD (26.47)} = 1.05$$

*Paired *t*-test pretest/posttest *p* = .010

**Paired *t*-test pretest/posttest *p* = .654