Evaluation of Two Interviewing Skills Measures:
An Instrument Validation Study

Cathy King Pike
Robert Bennett
Valerie Chang

Abstract: *This article reports an initial validation of an instrument that measures basic interviewing skills and compares its psychometric results with another instrument that has been used more frequently to measure similar skills. Four field supervisors rated 30 students' videotaped interviews (*N=120*) using two instruments, the validation, and a comparison instrument. The current validation instrument had high internal consistency reliability, a clear factor structure, and performed well in construct validity evaluations. These preliminary results supported the instrument's internal consistency reliability, content, factorial, and construct validity. The validation instrument had higher internal consistency reliability, lower error measurement, and a more interpretable factor structure than the comparison instrument.*

Keywords: *Assessment; interviewing skills; instrument development; direct
measures; measurement*

Before beginning to work with clients, all social work students need to master basic practice skills. These basic skills are generic prerequisites to additional skills required in specialized fields of practice and for particular theoretical approaches. These basic skills are widely recognized as beginning, exploring, and contracting with clients (Hepworth, Rooney & Larsen, 1997).

Students are expected to learn basic skills in practice courses, later developing more complex competencies during field practice. In a study completed by Dore, Epstein and Herrerias (1992), eight field-training objectives were identified. Their first objective was the "development of specific skills for micro practice, including skills in engagement, problem exploration, exploration of feelings, goal setting, contracting, and termination, as well as knowledge of and ability to apply various treatment modalities" (Dore, et al., p. 357). Also noted by these authors was the paucity of student learning measures.

Learning basic skills and engaging in self-assessment contribute to becoming self-reflective social workers with the skills to continuously improve prac-

Cathy King Pike is Associate Professor, Robert Bennett is Assistant Professor, and Valerie Chang is Professor at the Indiana University School of Social Work, Indianapolis, IN 46202.

tice (Bloom, Fischer & Orme, 1995). Although students should be responsible for assessing their progress (Shepard & Wahle, 1981), they also need ongoing feedback from their classroom and field instructors (Stoltenberg & Delworth, 1987). Beginning students are particularly dependent upon supervisors for direction, feedback, and evaluation (Stoltenberg & Delworth, 1987).

Unfortunately, students report that they receive minimal feedback (Barth & Gambrill, 1984) and that they need critical analyses of their use of social work skills (Urbanowski, 1988). Relevant, practical, and psychometrically sound evaluation tools that can be used across practice settings to provide feedback to students about their interviewing skills are needed (Vourlekis, Bembry, Hall & Rosenblum, 1996).

With the increased pressure for accountability in practice and education (Bernotavicz, 1994), social work educators need outcome measures that accurately assess competency in the use of basic and more complex practice skills (Matarazzo & Patterson, 1986; O'Hare & Collins, 1997; Ragg & Mertlich, 1999). However, sound, psychometrically tested measurement tools for evaluating basic practice skills are not readily available (O'Hare & Collins, 1997; Vourlekis, Bembry, Hall & Rosenblum, 1992). Without solid evaluation tools, it is difficult to effectively and consistently evaluate beginning practice skills. Evaluation instruments that have good levels of reliability and validity can enhance the learning experience for students. Specifically, evaluation instruments should assess accurately students' skill development and identify skills that students need to develop further.

Only three instruments related to basic practice skills were identified in the social work literature: a measurement tool used in evaluating students' field process recordings (Vourlekis, et al., 1996), a social work practice skills instrument (O'Hare & Collins, 1998), and a somewhat dated interview skills assessment instrument (Katz, 1979). Vourlekis, et al. (1996) reported research on the usefulness of a checklist in evaluating interviewing skills in field. The instrument contains 26 items and is scored from "1=beginning level" to "5=advanced level." Vourlekis, et al. (1996) found high internal consistency reliability and good validity results for the instrument. However, this instrument was developed for use in evaluating apparent interviewing skills through the use of process recordings and is limited to students' self-report and the verbal content reported in a process recording. The checklist, although useful, is an indirect measure, rather than a direct measure of interviewing skills.

O'Hare and Collins (1997) reported the most recent instrument development project related to practice skills measurement. Their instrument addressed the frequency with which practitioners use 23 skills that are therapeutic, supportive, case management, and evaluation. The items are measured from 1 to 5, and the anchors range from "never/almost never" to "very often." This is a self-report instrument that better relates to the frequency with which MSW students and MSW practitioners (O'Hare & Collins, 1998) use a variety of skills in their practices, rather than measuring competencies in the use of these skills. In psychometric analyses, the instrument yielded

acceptable to strong internal consistency reliability coefficients (.78 to .92), with the four-factor solution remaining stable across two studies (O'Hare & Collins, 1997; 1998). Although this instrument is easy to use and can be completed quickly, the measurement goals, i.e., identifying the frequency of use of skills, differ from the instrument tested in this study.

The Katz (1979) instrument is more dated than the two instruments described above and has a complicated scoring mechanism. However, this instrument is similar to the current validation instrument and has been used more frequently in research on students' levels of interviewing skills, as reported in the Katz (1979) writing. The Katz instrument contains 23 items and was designed to include identifiable skills taught in beginning social work practice methods courses. The first 11 items on this instrument are descriptions of interviewer characteristics. The items are "attentive to clients," "eye contact," "relaxed," "self-conscious," "fidgety," "distracted," "genuine," "respect for client," "sensitive to client's feelings," "mutuality" and "warmth" (Katz, 1979). These items are rated using a four-point Likert-type scale ranging from "0=almost never present" to "3=almost always present."

Items 12 through 22 are associated with discrete interviewing skills. These skills are "verbal following," "exploratory responses," "understanding responses," "primary-level empathy," "summarizing responses," "self-disclosure," "advice," "confrontation," "advanced-level empathy," "immediacy," and "concreteness." For these items, the rater uses a three-point scale for appropriateness of use. The scale ratings are "0=not appropriate and not used," "1=over-use or under-use," "2=appropriate use." That score is then multiplied by a weight from a four-point scale of effectiveness with high effectiveness rated as a "4" and low effectiveness rated as a "1."

The final item, 23, requires an overall judgment about the student's competence as an interviewer when compared with other students. This item is rated on a seven-point Likert-type scale of "1=poor" to "7=excellent."

In preliminary studies conducted across three cohorts, inter-rater reliability ranged from poor to good, .48 to .82 (Katz, 1979). Katz' (1979) validation studies exhibited two major flaws in analyses. First was the use of 22 items in three very small samples (N=15 or fewer for the first two samples, and N=26 in the third sample) to predict overall performance, a global item contained in the instrument. These sample sizes were too small to provide stable multiple regression results (Pedhazur, 1973). Second, the Katz (1979) studies used an excessive number of dependent samples t-tests that were computed to examine changes from pretest to posttest. An astounding 23 analyses were conducted for each of the items of the scale (Katz, 1979). Cohen and Cohen (1983), using tables that take sample size into account, reported the estimated Type I error rate for 20 separate tests as being about 90%.

The objectives of this research were to: (1) psychometrically test an instrument designed by Chang and Scott (1999) but for which no psychometric testing had been completed, and (2) compare the Chang and Scott (1999) instrument's psychometric characteristics to those of the Katz (1979) instrument,

which has been more frequently used in research on basic interviewing skills. The Katz (1979) instrument was designed to offer instructors a tool for evaluating skills taught in the classroom and practiced in a laboratory setting. Although the Katz (1979) instrument does not contain some currently recognized practice skills, the instrument is a core set of interviewing skills items that are similar to those in the Chang and Scott (1999) instrument. Permission was obtained to include the Katz (personal communication) instrument in this research in order to compare the Chang and Scott (1999) instrument results to the more frequently used Katz (1979) instrument.

METHODS

Description of the Validation Instrument

The Chang and Scott (1999) instrument focuses on social workers' behaviors related to interviewing and is constructed with the goals of having faculty, field instructors, and students use it to evaluate students' interviewing skills. The instrument is designed to measure a variety of interviewing skills: communicating involvement, observing, active listening, beginning process, reflective questioning, exploration, seeking clarification, initial contracting, and interpersonal skills. See the appendix for additional details about the skills and behaviors or descriptors that comprise the broader level skills included in this instrument.

Within the first nine skills are lists of behaviors that are inherent to broader-level interviewing skills. For example, communicating involvement includes the following discrete behaviors: attentive body posture, facial expressions, and eye contact. The purpose in listing behaviors was to focus raters on the specific grouping of behaviors that comprise each overall skill. Each of these behaviors or descriptors is rated dichotomously (present or not present). The process of reviewing the dichotomously scored skills can assist raters in more accurately appraising the broader category. Similarly, students' can further their understanding of specific behaviors that will improve their interviewing competencies. For purposes of this study, these discrete behaviors, which comprise a broader skill, were not included in psychometric analyses. Each broader-level skill is rated on a five-point Likert-type scale (1 = ineffective and/or inappropriate to 5 = highly effective and appropriate).

An additional four items (10–13) focus on the interpersonal qualities of warmth, respect, empathy, and genuineness. Each of these final four qualities is more abstract than many of the practice skills, involves less universally recognized behaviors, and requires raters to make a somewhat subjective judgment. Because of this, no attempt was made to include specific lists of behaviors for these four items.

The final item (14) is focused on rating the general effectiveness of students' responses to clients. Scoring the instrument is accomplished by summing up the first 13 ratings. This item was included for purposes of using it in evaluating the instrument.

Sample and Informed Consent Procedures

Study participants were graduate social work students recruited from two sections of a first-year MSW theory and practice methods course. The course

objectives, textbooks, workbook, and required final videotape were the same in each section. An "Informed Consent" form, provided to students, described the following: The intent of the study, the students' extent of involvement, any risks and benefits associated with the study, information about the voluntary nature of the study, students' right to refuse to participate, and their right to withdraw their consent to participate in the study at any time during their involvement in the study. Thirty of 44 students (response rate = 68.2%) across the two course sections voluntarily participated in the study.

Demographic information was gathered at the time the students agreed to be involved in the study. The sample consisted of 26 females and four males. There were 27 Caucasian students, two African-American students, and one Hispanic student. The students ranged in age from 22 to 53 years old, with a mean age of 28 and a median age of 25 years.

Design and Procedures

Students in each class were videotaped conducting a 15-minute interview with an individual trained to simulate a single client. The authors hired two senior-level BSW students and two-second year MSW students to simulate the client. These students were included in the development of a client profile and practiced the client roles with coaching from the authors.

Four social workers with at least 10 years of post-MSW practice experience were recruited and trained to rate the videotapes. Their years of social work experience ranged from 10 to 23 years, with a mean of 14 years. All the raters were female and had from one to six years of field instructor experience. Three of the raters were Caucasian and one African American.

In a four-hour training session, the raters were instructed in the use of the Chang and Scott (1999) instrument, as well as the Katz (1979) comparison instrument. During the training, the raters watched two videotaped inter-views not included in the study and evaluated them using both the validation instrument and the Katz (1979) instrument. Raters' evaluations were compared and discussed by the trainers and raters.

Each rater received copies of the 30 student videotapes and evaluation instruments. The raters were instructed to evaluate each tape using both the revised Chang and Scott (1999) and the Katz (1979) instruments (N=120) and complete and return the ratings within two months.

Psychometric Examinations

Several statistical analyses examined the psychometric properties of this instrument in relation to the Katz (1979) instrument. Internal consistency reliability evaluations were conducted through computations of Cronbach's coefficient alpha. Information about inter-rater reliability was obtained through Analyses of Variance (ANOVA) and Tukey's HSD follow-up tests. When the important assumption of homogeneity of variance was violated by one data set, Kruskal-Wallis tests (the non-parametric analogue to ANOVA) were used.

Principal Components factor analyses using Promax rotations examined the content, construct, and factorial validity of the instrument. These factor

analyses were interpreted through the use of the percentage of variance accounted for by factors, eigenvalues, visual inspection of the Scree plot, and theoretical considerations.

A final analysis to examine evidence of construct validity was conducted by computing a Pearson's correlation for the Chang and Scott (1999) instrument scores with the Katz (1979) instrument scores. The purpose of this examination was to test the hypothesis that the two instruments would have a positive, moderate correlation, suggesting that the two instruments are measuring related, but different, constructs.

No attempt was made to use individual items to predict overall ratings (item 14) of the students' responses to clients, because the sample was too small to permit the large number of predictors (13) that would be needed for computing a multiple regression analysis. However, a Pearson's Product-Moment correlation was computed for the Chang and Scott (1999) summed scores (items 1-13) and the global rating item (item 14) to assess their correlation and the amount of variance in the global rating scores that can be accounted for by the 13 items taken together. Item 14 asks the rater to provide an overall effectiveness of students' responses to clients. It is scored 1-5, from "1=Ineffective" to "5=Highly effective."

RESULTS

Descriptive Statistics and Initial Examinations

Descriptive statistics were generated for both the Katz (1979) and the current version of the Chang and Scott (1999) scale. The Katz (1979) scale scores ranged from 45 to 127 (Mean = 78.62, Mdn = 80.00, Std. Dev. = 17.68). The Chang and Scott (1999) instrument's scores ranged from 31 to 76 (Mean = 58.18, Mdn = 58.0, Std. Dev. = 8.68). A rater scored one videotaped interview on the validation instrument as 31, which was more than two standard deviations below the mean. Because outliers like this score can unduly influence small samples, this case was deleted from the database (N=119). The data were examined for additional positive outliers, but none were identified within the analyses as highly atypical.

There were some differences between the two instruments on the issue of missing data. The Katz (1979) scale had no missing data. However, the Chang and Scott (1999) instrument had enough missing data to bring its sample size to 86 (rather than 119) in reliability analyses where listwise deletion is used. Frequencies were generated for individual items to assess systematically the extent to which data were missing. Most of the items in the Chang and Scott (1999) instrument had minimal missing data (ranging from 0-6). Two items (8 and 9), respectively, had 19 and 14 missing data-points. Item 8 referred to seeking clarification. The dichotomous statements used to focus raters were as follows: exploring the meaning of clients' words, conclusions, contradictory statements, and eliciting detail about statements. Item 9 involved the contracting process and, likewise, seemed well defined through the use of its dichotomous focusing statements. Those items where one might expect missing data because of their level of abstraction (e.g., interpersonal qualities of warmth, respect, empathy, and

genuineness) had no missing data. No explanations were apparent in these examinations, which explained the missing data. The sample sizes for internal consistency reliability analyses, by definition, deleted cases where data were missing (listwise deletion). However, factor analyses accommodated the missing data through the use of a pairwise deletion.

Evaluations of Reliability

Internal Consistency

Cronbach's coefficient alphas were computed for both the Katz (1979) and the current version of the Chang and Scott (1999) scales. Analyses were computed with and without the global rating item of each scale. The Katz (1979) scale achieved an acceptable internal consistency reliability score for research purposes, .77 (N=119) for all items and a .74 when the global rating item was removed from analysis.

The Chang and Scott (1999) scale achieved an excellent level of reliability of .91 (N=86) for all items and a closely matching .90 when the global item was removed from analysis. The Chang and Scott (1999) instrument had substantially lower standard errors of measurement than the Katz (1979), indicating that the Chang and Scott instrument (1999) exhibits more precise measurement of basic interviewing skills than the Katz (1979) instrument. Table 1 contains further information about the internal consistency reliability examinations.

Figure 1: *Internal Consistency Reliability Results*			
Instrument Evaluated	**Std. Dev.**	**Alpha**	**SEM**
Katz (1979)–all items	15.17	.77	7.35
Katz (1979)–global item omitted	14.28	.74	7.13
Chang & Scott (1999)–all items	7.39	.91	2.04
Chang & Scott (1999) –global item omitted	6.74	.90	2.16
Using Factor Analysis Results			
Katz (1979)			
Factor 1	10.82	.81	4.70
Factor 2	7.88	.54	5.32
Chang & Scott (1999)			
Factor 1	6.48	.91	1.91
Factor 2	1.35	.61	.84
Note: Results are rounded to two places.			

Inter-rater Reliability

Inter-rater reliability was assessed for both instruments via Analyses of Variance (ANOVA), where rater was the independent variable and scale score comprised the dependent variable. The Levene's test for the Katz (1979) data resulted in a significant F-score (F=11.034, p.<.001), indicating that the homogeneity of variance

assumption of the ANOVA had been violated. Consequently, the non-parametric Kruskal-Wallis was computed for these data. It yielded a non-significant Chi-Square of 5.67 ($p.$=.129), indicating no significant differences by rater on mean rank (ranging from a low of 47.48 for rater 3 to 65.85 for rater 1).

The data for the revised Chang and Scott instrument (1999) did not violate the homogeneity of variance assumption by the Levene's F-test (F=.813, $p.$=.489). The overall ANOVA indicated significant differences in mean ratings among the raters (F=9.137, $p.$<.001). The Tukey's HSD follow-up test indicated that rater 3 had a significantly lower mean rating (52.37) than the remaining three raters (ranging from 58.40 to 61.57).

Frequencies were generated for all raters in order to further understand the differences among raters' distributions. Hand computations of t-tests for kurtosis and skewness for all raters were not significant (df=29, tskewness ranged from .115 for rater 4 to -.806 for rater 2, and tkurtosis ranged from -.278 for rater 2 to –1.358 for rater 4). However, an examination of the frequency distributions indicated that fully half of rater 3's ratings were below a score of 50, compared to the frequencies of 1, 3, and 3 scores that were below 50 for raters 1, 2, and 4, respectively. Median ratings were 51 for rater 3 and 61, 62, and 56 for raters 1, 2, and 4, respectively.

Based upon the above obvious inconsistency of the third rater's ratings, the Kruskal-Wallis and ANOVA were recomputed. In this computation, rater 3's ratings were removed to identify the extent to which this person's ratings were unduly influencing the overall inter-rater reliability. Rater 3's data were retained for all other analyses, however. The Katz (1979) data again violated the homogeneity of variance assumption (Levene's F=16.050, $p.$<.001), and the subsequently computed Kruskal-Wallis Chi-Square was not significant (.148, $p.$=.929).

The Chang and Scott instrument (1999) data again did not violate the homogeneity of variance assumption (Levene's F=1.206, $p.$=.304). The ANOVA for these data yielded a non-significant F of 1.058 ($p.$=.352, Eta-Square = .02). Only about 2% of the variance in ratings can be attributed to rater differences when rater 3's data was removed from the analyses.

The mean ranks and means across both data sets were similar for raters 1, 2, and 4. The Katz instrument (1979) mean ranks ranged from 43.53 to 45.88, and the Chang and Scott instrument (1999) means ranged from 58.83 to 61.57, indicating good evidence of inter-rater consistency across these three raters in scoring both instruments. Rater 3's inconsistent ratings call somewhat into question the inter-rater reliability for the Chang and Scott (1999) instrument. However, the results, after removing Rater 3's data, are more suggestive of rater, rather than instrument, inconsistency. This finding is based on both the insignificant results and the fact that when Rater 3's data were removed from the sample, only 2% of variance among the remaining raters can be attributed to rater differences. However, Rater 3's data were included for all other analyses, including evaluations of internal consistency reliability, factor analyses, and all other validity examinations. This decision prevents the loss of data but also provides more conservative psychometric estimates.

EXAMINATIONS OF FACTORIAL, CONTENT, AND CONSTRUCT VALIDITY

Factor Analysis

Principal components factor analyses were computed to examine the factor structure of both the revised Chang and Scott (1999) and Katz (1979) instruments. The Kaiser-Meyer measure of sampling adequacy for the Katz (1979) items yielded a score of .868 and a Bartlett's Test of Sphericity Chi-Square of 1,450.92 (p.<.001). These two tests, respectively, examine the extent to which the items in the analysis are sufficiently correlated and differ significantly from an identity matrix. Both tests indicated that the items were well suited for examinations using factor analyses. Likewise, these two tests of assumptions were met with the Chang and Scott (1999) items, yielding a Kaiser-Meyer coefficient of .923 and a Bartlett's Chi-Square of 710.59 (p.<.001).

The initial factor analysis of the Katz (1979) instrument yielded a 4-factor solution, with the factors accounting for the following percentage of variance, respectively: 36.84, 11.29, 6.49, and 5.62. An examination of the Scree plot indicated that the last two factors were likely comprised of error variance. The first two factors were correlated at .484, and factors 3 and 4 had low to trivial correlations with all factors. The correlations of the last two factors with the other factors ranged from -.000006 for factors 3 and 4 to .377 for factors 3 and 2. Factor 4 was negatively correlated with all other factors, with its highest correlation being a -.329 with the first factor.

A second factor analysis was conducted with the Katz (1979) items, where a 2-factor solution was specified. In this analysis, 36.84% of the variance was attributed to factor 1 and 11.29% to factor 2, for a total of 48.13% of variance accounted for by the two factors. All but six items loaded most heavily on the first factor. Those items loading most heavily on factor 2 were items 17-22. These items were designed to measure self-disclosure, the provision of advice, use of confrontation, "advanced-level" empathy (as opposed to "primary-level" empathy), discussion of current therapeutic relationships, and concreteness. Factors 1 and 2 had a very low correlation (r=.225), indicating that these factors measure different latent constructs and should be scored separately.

The initial factor analysis of the revised Chang and Scott (1999) instrument indicated the presence of two factors with 50.88% of the variance accounted for by the first factor and 8.71% for the second factor. Only three items loaded most highly on factor 2 (items 5-7). These items related to assessing client problems in relation to: (1) the nature of the problem, its history, severity, and precipitating factors; (2) the problem's effects on the person's feelings and functioning, and the client's personal strengths; and, (3) situational stresses, supports, and strengths. These three items are highly correlated and more focused than items stated at a general level, e.g., general interpersonal, process, and exploring skills. However, they are central components of basic interviewing skills in social work. Further, the two factors were moderately correlated at .54, a level that Nunnally and Bernstein (1994) considered sufficiently high to combine factors.

A second factor analysis was conducted specifying a one-factor solution. The total percentage of variance explained by this solution was 50.80%. In this analy-

sis, items 5, 6, and 7 loaded positively on the single factor, with a range from .372 (item 6) to .606 (item 5).

Internal Consistency Reliability of Identified Factors

Based on the findings of the above factor analyses, changes to both the Katz (1979) and the Chang and Scott (1999) instruments' internal consistency reliability were re-examined. Two evaluations of the internal consistency reliability were conducted for each factor of each instrument using Cronbach's coefficient alpha. First, internal consistency reliability analyses were computed on the Katz (1979) instrument for the two factors identified through factor analyses. The internal consistency reliability and the standard error of measurement (SEM) of Factor 1 were improved through deletion of items 17-22 (alpha=.81, SEM=4.70), but the internal consistency reliability still remained lower and the SEM higher than had been found for the Chang and Scott instrument (1999) in earlier analyses. In addition, the internal consistency reliability coefficient for Factor 2 was unacceptably low (alpha=.54), and the SEM unacceptably high (SEM=5.32) for a scale having only six items and a narrow possible range of scores. The items identified as comprising factor 2 are not internally consistent and exhibit a high level of measurement error.

Internal consistency reliability analyses were computed on the revised Chang and Scott (1999) instrument for the two factors initially identified by the factor analysis. The first factor's internal consistency and SEM remained essentially unchanged (alpha=.91, SEM=1.915). The second factor, comprised of only three items, had an unacceptably low level of internal consistency reliability (alpha=.61) for evaluating individual interviewing skills, but its SEM (SEM=.844) also was very low in relation to the possible range of scores (possible range = 3-15). Scales containing only three items tend to suffer from low levels of internal consistency reliability and poorly represent the breadth of constructs (Nunnally & Bernstein, 1994). For these reasons, the three items were retained with the other items to comprise a single scale. Further, when these three items were removed from the scale, their omission did not change an already very strong level of internal consistency reliability, and the items' inclusion with the other items yielded only a very trivial amount of increase in error measurement. From a theoretical framework, the inclusion of the three items is sound, because those skills measured by the three items are central aspects of conducting effective interviews.

Further Evaluation of Construct Validity

The final analyses evaluated the construct validity of the Chang and Scott (1999) instrument (see Table 2). A Pearson's Product-Moment correlation was computed for the Chang and Scott (1999) revised instrument scores and the older, more frequently used Katz (1979) instrument scores. Based on the similarity of some of the items in both scales, a positive moderate correlation coefficient had been hypothesized. The Pearson's correlation resulted in a positive and moderate correlation between the two instruments (r=.585, p.<.001). This result supported the hypothesis that the two instruments would measure somewhat different but similar constructs.

Figure 2: *Final Factor Analyses Results*

Validation Instrument		Katz (1979) Instrument		
Items	One-Factor	Items	Factor 1	Factor 2
			(Structure Matrix)	
13.	.87	10.	.80	.22
12.	.85	23.	.80	.54
14.	.84	9.	.76	.15
10.	.81	6.	-.76	-.09
11.	.80	11.	.75	.12
1.	.80	5.	-.73	-.08
4.	.77	2.	.72	-.23
8.	.75	3.	.70	.23
3.	.72	13.	.70	.46
5.	.61	8.	.68	-.17
2.	.60	7.	.67	.06
9.	.53	1.	.67	-.27
7.	.44	12.	.68	.39
6.	.37	14.	.66	.49
		15.	.6	.43
		4.	-.59	-.28
		16.	.54	.48
		21.	.17	.54
		18.	.09	.54
		17.	-.04	.50
		20.	.03	.50
		22.	.14	.41
		19.	.08	.38

Note: Items are rounded to two places.

The Pearson's Product-Moment correlation between the summed Chang and Scott (1999) instrument scores (items 1-13) and the global rating item (item 14) resulted in a positive, moderate, and significant relationship (r=.676, p.<.001, N=119). About 46% of the variance in the global item can be accounted for by scores on the first 13 items. This solid relationship between the global item that measures overall effectiveness of student responses to clients with the individual items taken together provides preliminary evidence of construct validity.

DISCUSSION AND APPLICATIONS TO SOCIAL WORK PRACTICES

Like much research that is based in field or natural settings, i.e., classrooms, this research has several limitations. First, the research used a single sample of social work students from one school and across sections of only one course. Further research is needed to examine the instrument's psychometric properties with more diverse groups of social work students and at varying academic levels and courses.

Second, the Katz (1979) instrument was not ideal as a comparison instrument, because it is dated and was examined using problematic validation techniques. Therefore, any interpretations based solely on this instrument require caution. However, there were strengths which using the Katz (1979) instrument provided the research. The Katz (1979) measurement purpose matched the purpose of this research, and the Katz (1979) instrument shared a common set of core interviewing skills with the Chang and Scott (1999) instrument. In addition, by examining the internal consistency reliability and factorial validity of the Katz (1979) instrument, preliminary information not only was obtained regarding the Katz (1979) instrument's psychometric properties, but it allowed a comparison between the Katz (1979) instrument and one (Chang & Scott, 1999) that had not previously been psychometrically evaluated. These analyses were positive steps toward mediating the limitations of using an instrument with unknown psychometric qualities.

A third limitation of the study was substantial missing data for two items of the Chang and Scott (1999) instrument compared to no missing data in the Katz (1979) instrument. These were the two items that measured seeking clarification and contracting process skills. It is unclear whether the missing data are due to some aspect of formatting or a lack of clarity in the items. However, the Cronbach's alpha analysis indicated that both items had moderate to strong corrected item-total correlations (.70 and .46, respectively), indicating that they were internally consistent, i.e., contributed to accurate measurement.

A fourth limitation became evident in examinations of inter-rater reliability, specifically those of rater 3, whose scores were substantially different from the other raters. With this research, it was not possible to identify whether the differences in this rater's scores were due to the rater's incomplete training, the researchers allowing too much time for raters to complete the rating tasks, the rater's procrastination in completing the ratings, or unreliability in the instrument. Further research should be conducted on inter-rater reliability. However, other findings from this study suggest that instrument unreliability is not likely the cause of rater 3's inconsistent ratings. Specifically, the strong results on virtually every psychometric evaluation argue against instrument unreliability as the explanation for rater 3's inconsistency in rating.

A fifth limitation is that this research used simulated clients and, because of that, it is impossible to know how valid and reliable this instrument would be when applied to interviews with actual clients. Further research is needed to examine the usefulness of the instrument with actual clients.

Despite the preliminary nature of findings from this study, the Chang and Scott (1999) instrument performed very well in evaluations of its internal consistency reliability, inter-rater reliability (when rater 3's scores were omitted for this analysis), and content, factorial, and construct validity. Both the evaluations of internal consistency reliability, where all 14 items were used, and the evaluation with the global item omitted, yielded high enough alpha coefficients for the instrument to be used in evaluating individual students' basic interviewing skills. Furthermore, the instrument had higher internal consistency reliability coefficients when its items were used as hypothesized, rather than those computed on the basis of the

factor analysis results. The factor analyses provided evidence of good content, factorial, and construct validity. The correlation analysis also provided preliminary evidence of construct validity, resulting in a positive, moderate correlation with another instrument that has been used to measure interviewing skills.

In contrast, the Katz (1979) instrument fared less well on virtually all analyses than the Chang and Scott (1999) instrument. The Katz (1979) instrument had substantially lower alpha coefficients and higher measurement error on both analyses, where all of its items were analyzed, and then with the global item measuring interviewing skills omitted. An examination of this instrument's factorial structure found two factors, rather than the implied one-factor, solution. Further analyses of internal consistency based on the factor analyses supported a two-factor solution for the instrument, substantially increasing the level of the alpha coefficient for the majority of items when the items comprising the second factor were omitted from the reliability analysis. The Katz (1979) instrument and the Chang and Scott (1999) instruments had a moderate positive correlation, providing very preliminary evidence from the Katz (1979) instrument of convergent construct validity for the Chang and Scott (1999) instrument.

A larger validation study that includes a representative sample of students from a diverse sample of schools could provide additional information about the instrument's usefulness as a tool in evaluating interviewing skills. Increased structure and data collection controls may improve the instrument's inter-rater reliability. In addition, further research is needed to examine the extent to which basic interviewing skills are transferred from class exercises to work with clients in field placements.

From a teaching viewpoint, an instrument that exhibits good evidence of reliability and validity and which can be used in both the classroom and the field may be helpful in coordinating learning across the areas. A further advantage would be the ability to promote discussions across the two learning environments regarding learning needs and challenges. Perhaps such consistency would promote the better transfer of learning from the classroom to field settings and, ultimately, to effective social work practice. This research sought to achieve a preliminary step toward facilitating coordination in learning across the classroom and the field by psychometrically evaluating two instruments that evaluate students' basic interviewing skills.

References

Barth, R., & Gambrill, E. (1984). Learning to interview: The quality of training opportunities. *The Clinical Supervisor, 2*, 3-14.

Bernotavicz, F. (1994). A new paradigm for competency-based training. *Journal of Continuing Social Work Education, 6*, 3-9.

Bloom, M., Fischer, J., & Orme, J. (1995). *Evaluating practice: Guidelines for the accountable professional.* Boston: Allyn & Bacon.

Chang, V.N., & Scott, S.T. (1999). *Basic interviewing skills: A workbook for practitioners.* Chicago, IL: Nelson-Hall.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dore, M.M., Epstein, B.N., & Herrerias, C. (1992). *Evaluating students micro practice field performance: Do universal learning objectives exist? Journal of Social Work Education, 28*(3), 353-362.

Hepworth, D.H., Rooney, R.H., & Larsen, J.A. (1997). *Direct social work practice: Theory and skills* (6th ed.). Pacific Grove, CA: Brooks/Cole.

Katz, D. (1979). Laboratory training to enhance interviewing skills. In F.W. Clark & M.L. Arkava (Eds.), *The pursuit of competence in social work* (pp. 205-226). San Francisco: Jossey-Bass.

Matarazzo, R.G., & Patterson, D.R. (1986). Methods of teaching therapeutic skill. In S.L. Garfield & A.E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 821-843). New York: John Wiley & Sons.

Nunnally, J., & Bernstein, I.H. (1994). *Psychometric theory* (3rd Ed.). New York: McGraw-Hill.

O'Hare, T., & Collins, P. (1997). Development and validation of a scale for measuring social work practice skills. *Research on Social Work Practice, 7*(2), 228-238.

O'Hare, T., & Collins, P. (1998). Validation of the Practice Skills Inventory with experienced clinical social workers. *Research on Social Work Practice, 8*(5), 552- 564.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research* (2nd Ed.). Fort Worth: Holt, Rinehart and Winston.

Ragg, M.D., & Mertlich, G. (1999). Toward measuring practice skill outcomes: Three measures of practice skill. Paper presented at the Annual Planning Meeting of the Council of Social Work Education, San Francisco, CA.

Shepard, G., & Wahle, L.P. (1981). A competency-based approach to social work education: Does it work? *Journal of Education for Social Work, 17*(3), 75-82.

Stoltenberg, C.D., & Delworth, U. (1987). *Supervising counselors and therapists: A developmental approach*. San Francisco: Jossey-Bass.

Urbanowski, M.L. (1988). *Learning through field instruction: A guide for teachers and students*. Milwaukee, WI: Family Service of America.

Vourlekis, B., Bembry, J., Hall, G., & Rosenblum, P. (1992). Evaluating the Interrater Reliability of Process Recordings. *Research on Social Work Practice, 2*(2), 198-206.

Vourlekis, B., Bembry, J., Hall, G., & Rosenblum, P. (1996). Testing the reliability and validity of an interviewing skills evaluation tool for use in practicum. *Research on Social Work Practice, 6*(4), 492-503.

Author's Note:

Address correspondence to: Valerie Chang, Professor, Indiana University School of Social Work, 902 West New York Street, ES 4153, Indianapolis, IN 46202, USA. E-mail: vchang@iupui.edu

Appendix

Definitions of Skills in the Chang and Scott (1999) Instrument

1. Communicating involvement includes attentive body posture, facial expressions, and eye contact.

2. Beginning process skills consist of the ability to communicate beginning information such as describing the initial purpose of the meeting, explaining the process, discussing ethical and agency policies and introducing yourself and your role.

3. Questioning skills involve balanced use of open-ended and close-ended questions.

4. Exploring problem/challenge contains such areas as gaining information about previous attempts to solve the problem, history of the problem, and severity of the problem.

5. Exploring person includes seeking information about feelings about having the problem, effects of the problem on functioning, and personal strengths.

6. Exploring the situation consists of gaining information about effect of the problem on others, available social support, other demands and stresses in the situation/environment, and strengths in the situation/environment.

7. Reflecting skills range from the ability to summarize the client's feelings and basic content to the ability to encapsulate themes related to behavior, thoughts, feelings, and interactions with others.

8. Seeking clarification skills entails using questions to explore areas such as the meaning of words, the basis of conclusions, statements that appear contradictory, and details about sequences of interaction.

9. Contracting process skills range from the ability to reach agreement about problems to establishing clearly defined goals and creating a contract.

10. Expressing warmth involves verbal and nonverbal expressions of concern and compassion.

11. Expressing respect is defined as communicating regard for such things as the client's feelings, thoughts, potential, strengths, and resources.

12. Expressing empathy is defined as communicating understanding and acceptance of the client's felt experience.

13. Expressing genuineness includes being sincere, fully present, and able to able share reactions with the client.

14. Effectiveness of responses involves using interventions that invite the client into further exploration.