Bilingual computerized speech-recognition screening for clinical depression: Evaluating a cellular telephone prototype

GERARDO M. GONZÁLEZ, CRAIG R. COSTELLO, MARIO VALENZUELA,
BEVERLY CHAIDEZ, and ARCELA NUÑEZ-ALVAREZ
California State University, San Marcos, California

This exploratory field study evaluated a bilingual computerized speech-recognition cellular telephone prototype of the Center for Epidemiological Studies—Depression scale (CES-D). Thirty Spanish and 22 English speakers completed both computer-telephone and face-to-face CES-D methods and an oral depression checklist in counterbalanced order. Both language groups reported high positive ratings for the computer-telephone method, with the English sample preferring the computer-telephone over the face-to-face method. In both samples, the computer-telephone method yielded high internal consistency estimates, strong alternate form reliabilities, and similar high correlations to the depression checklist. Both groups reported significantly elevated scores with the computer-telephone method, but total score variances for both methods did not differ. Computer-telephone limitations included occasional misrecognitions and template training constraints.

Among the most critical national public health concerns is clinical depression. Between 10% and 25% of the general population report significantly high depressive symptoms during any 1-month period (Robins et al. 1985; Weissman, Bruce, Leaf, Florio, & Holzer, 1991). In addition, the estimated direct and indirect economic costs of depression increased from 16 billion to 43 billion dollars during the past decade (Greenberg, Stiglin, Finelstein, & Berndt, 1993; Stoudemire, Frank, Kamlet, & Hedemark, 1987). About 75% of clinically depressed persons in the general population initially seek a health-care provider rather than a mental-health professional for treatment (Shapiro, et al. 1984). In comparison, only 11% of Mexican Americans (relative to 22% of non-Hispanic whites) who met the criteria for clinical depression sought mental-health professionals (Hough et al. 1987). As many as 30% of patients in primary-care settings report significant depressive symptoms (Broadhead, Clapp-Channing, Finch, & Copeland, 1989). Primary-care clinics, however, generally have high patient volume and significant time constraints that hinder adequately assessing depression. In one study, Pérez-Stable, Miranda, Muñoz, and Ying (1990) found that over half of primary-care patients were misdiagnosed for depression, despite lenient criteria, by nonpsychiatric health-care personnel. Consequently, many high-risk and

actual cases of clinical depression remain undetected and untreated.

Computer-Assisted Assessment

Computer-assisted applications have offered strategies to facilitate such psychological services (Fowler, 1985) as depression assessment. Research has suggested that depressed patients report computerized interactive interviewing as acceptable or even preferable to human interviewing (Carr, Ghosh, & Ancill, 1983; Moore, Summer, & Bloor, 1984). In addition, depressed patients disclose their suicidality more often during computerized interviewing than during face-to-face interviewing (Levine, Ancill, & Roberts, 1989). Other studies have found that various computerized assessment methods were reliable and equivalent to conventional methods (Honaker, Harrell, & Buffaloe, 1988; Wilson, Genco, & Yager, 1985), including depression assessment (Kobak, Reynolds, Rosenfeld, & Greist, 1990).

A promising alternative to conventional interviewing techniques is computerized speech recognition. Computerized speech-recognition technology affords digital verbal presentation of discrete choice items, recognition of spoken responses, and scoring of the responses. Among the advantages of speaker-dependent speech recognition are more efficient, hands-free, real-time interaction in any language or accent (Bergeron, 1991). This technology offers potential assessment for persons not reliably assessed with English-language paper-and-pencil questionnaires, such as nonliterate individuals or monolingual non-English speakers (Starkweather & Muñoz, 1989). For example, computerized screening at primary-care settings may provide crucial information to nonpsychiatric health-care staff for appropriately referring patients to depression prevention or treatment (Muñoz, 1993; Muñoz & Ying, 1993).

The senior author acknowledges the support provided by a CSUSM Faculty Affirmative Action grant, CSUSM Center for Multicultural Studies, and a CSUSM Arts & Sciences Faculty Development grant for the research, development, and testing of the prototype. Thanks are also extended to John Copeland and Richard Serpe for their comments on the initial manuscript. All correspondence should be addressed to G. M. González, Psychology Program, California State University, San Marcos, CA 92096 (fax: 619-471-4156).

Thus, the capabilities of mental-health-care professionals may be enhanced, not substituted, with the aid of speech computerized tools.

Speech-Recognition Research

Several pioneering studies have successfully tested computerized speech-recognition psychological-assessment applications. Richards, Fine, Wilson, and Rogers (1983) developed a voice-recognition system for administering the Minnesota Multiphasic Personality Inventory (MMPI) to 32 disabled patients with limited hand function. The system visually displayed the MMPI items on a monitor, recognized the patient's verbal response, and generated a profile. The results indicated that there were no significant differences between the profiles produced by the computerized and paper-and-pencil methods.

Muñoz, González, and Starkweather (1991) pilot-tested an IBM-compatible speech-recognition "talking" prototype with 19 English- and 19 Spanish-speaking depressed primary-care medical patients. The program verbally presented the Center for Epidemiological Studies—Depression scale (CES-D), recognized the patient's oral responses, and generated a report of the patient's level of depressive symptoms. The results of the counterbalanced study suggested that speech-computerized and paper-and-pencil versions of the CES-D did not differ in total score means and variances and yielded high-reliability estimates for both samples. Moreover, English speakers displayed a preference for the computerized method.

González (1993a) developed a Macintosh speech-recognition "talking" CES-D prototype. A sample of 68 English-speaking participants completed computerized and paper-and-pencil forms of the CES-D and a computer anxiety scale in counterbalanced order. The results suggested that there were no significant differences between total score means and variances for the two CES-D methods. The two methods displayed high equivalent-forms reliability and internal consistency estimates. Moderate correlations between the CES-D methods and the computer anxiety scale were similar. Furthermore, participant preference rates of the CES-D methods did not differ (González, Spiteri, & Knowlton, 1995).

Telephone-Assisted Interviewing

Face-to-face interviewing is a conventional assessment technique (Andersen, 1993), but this approach is limited with Spanish-speaking communities because of language incompatibility, access constraints, or respondent suspicions of exploitation (Marín & Marín, 1991). An alternative technique is the telephone-assisted interview (Lavrakas, 1987). Marín, Pérez-Stable, and Marín (1989) found that telephone interviewing generated lower refusal rates for Latino participants than for non-Hispanic whites. Latino respondents perceived telephone interviews as personable and displayed greater willingness to answer highly sensitive questions on drug use and sexual behavior over the telephone than in a face-to-face situation (Marín & Marín, 1989).

An innovative data-gathering approach to increasing access for under-served populations is the cellular tele-

phone interview. Cunningham, Robinson, and Serpe (1993) interviewed homeless persons by cellular telephone to gather service-utilization data. The findings suggested that participant responses on the telephone and in face-to-face interviews were not significantly different. Based on the potential of cellular telephone interviewing, González (1993b) developed a speech-recognition cellular telephone prototype to screen for depressive symptoms among English and Spanish speakers.

Purpose of the Study

This exploratory field study evaluated an all-audio all-verbal speech-responsive computer program that administered a depression-screening questionnaire, via cellular telephone, to English- and Spanish-speaking samples. Our methodology and data analyses focused on the acceptability, administration times, and psychometric properties of the computer-telephone prototype.

Research Hypotheses

The research hypotheses included an evaluation of "equivalency" by comparing the computer-telephone prototype with a face-to-face version of the same depression measure.¹ Specifically, it was hypothesized that (1) respondents would report similar acceptance ratings and preference rates for the two methods, (2) the two screening methods would not differ in the total score means and variances, (3) the two screening methods would yield high alternate form and internal consistency reliability estimates, and (4) the two screening methods would display correlations similar to those of an independent depression measure.

METHOD

Sample

Initially, 36 Spanish- and 24 English-speaking adults, recruited from three health- and social-service facilities located in the San Diego area, completed the interviews. Eight participants were eliminated from the study, 7 because they did not reliably comprehend the computer-telephone instructions and 1 who, because of a 2-SD difference between CES-D scores, was considered a data outlier. The final sample consisted of 30 Spanish and 22 English speakers ($N = 52$). The Spanish-speaking group was 70% female and the English-speaking group was 54% female. Ninety-seven percent of the Spanish-speaking sample reported Latino ethnicity (83% identified as Mexican and 4% Nicaraguan; 13% declined to specify). Among the English-speaking sample, self-identified ethnicity was 82% white, 9% African American, and 9% other.

Participants ranged in age from 18 to 67 years. An independent-samples t test revealed a significant difference in the age levels of the two groups [$t(50) = 3.71, p < .001$, two-tailed]. Reported education levels ranged from 0 to 18 years for the entire sample. The sample variances for education were not equivalent [$F(1,50) = 3.33, p < .01$]. A separate variance estimate for an independent-samples t test indicated that these means were significantly different [$t(46.79) = 7.10, p < .001$, two-tailed]. The participants reported their computer experience using a 1–5 rating (1 = *no experience* and 5 = *very knowledgeable*). A t test revealed a significant difference in the reported computer experience of the two groups [$t(47) = 3.03, p < .005$, two-tailed]. Thus, the results suggested that the Spanish-speaking sample was younger, had fewer years of education, and reported less computer experience than did the English-speaking sample.² Table 1 summarizes the sample characteristics.

Procedures

The field study lasted from September to December 1993. Four (two bilingual) interviewers separately approached English- and Spanish-speaking adults at the field settings. Each interviewer explained the purpose of the study, clarified that participation was strictly voluntary and without any compensation, and obtained written consent. Participants completed the interview in their preferred language. Each participant responded to demographic questions and received instructions for completing three randomly assigned depression instruments. Each respondent completed both computer-telephone and face-to-face forms of a 20-item depression-screening measure in counterbalanced order. After individually completing each method, the interviewer recorded observed and expressed participant reactions to the method and an acceptance rating between 1 and 10 (1 = *very negative* and 10 = *very positive*). After completing both methods, the participant indicated a preference between the two methods and his/her reasons for the choice. Each participant also responded to an oral 16-item depression-symptom checklist. All data gathered during the interview were confidential and safeguarded. At the end of the session, the interviewer debriefed each participant.

Instruments

The Spanish-language instruments were utilized in established translated form or were appropriately translated by a bilingual expert.

The Center for Epidemiological Studies—Depression scale (CES-D). The 20-item self-report scale was designed to measure symptoms of depression (Radloff, 1977). One of four possible responses and associated weighted values (less than 1 day = 0, 1 to 2 days = 1, 3 to 4 days = 2, and 5 to 7 days = 3) indicated the frequency, during the previous week, of instances in which the respondent had felt as described in the statements. The CES-D included four reverse-scored items phrased in a nondepressive manner. The 20 weighted responses added to a total score that ranged from 0 to 60. Accordingly, a score of 16 or greater suggests a high level of symptoms of depression (Comstock & Helsing, 1976; Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977). The CES-D was selected for this study because of its strong reliability and validity and its widely established use with English- and Spanish-speaking populations (Mosciki, Locke, Rae, & Boyd, 1989; Roberts, 1980).

A depression symptom checklist. The orally administered 16-item depression measure, adapted from the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R) (American Psychiatric Association, 1987) criteria for major depression, employs a dichotomous response scheme (*No* = 0 and *Yes* = 1). To attain a total level of depression symptoms, the number of affirmative responses are summed. This checklist was selected as an independent validity criterion and has been used as a secondary depression measure with a number of English- and Spanish-speaking samples (S. A. Aguilar-Gaxiola, personal communication, April 12, 1994).

A structured demographic interview. The interview technique developed for this study comprised three sections. In the first, each participant provided demographic information elicited by the inter-

viewer (gender, ethnic identification, age, years of education, and computer experience). In the second, the respondent's reactions to administrations of the telephone and face-to-face CES-D methods were noted. In the third section, information regarding method preference and the respondent's primary reasons for the choice was obtained.

Computerized Speech-Recognition Telephone CES-D Prototype

The computer program utilized a Macintosh Centris 650 with 12-Mb random access memory, 240-Mb hard disk storage, and CD-ROM. The computer was connected to an ImageWriter printer, located at a secure university facility and accessible only to the research team. HyperCard 2.1 scripting (Claris Corporation, 1991) and the Voice Navigator speaker-dependent speech-recognition application (Articulate Systems, 1993) supported the program. Voice Navigator also linked the computer program to the telephone system. A Motorola TVS200 transportable cellular telephone with 3 W of power and up to 45 min of continuous talk time provided a portable microphone and speaker interface. Once activated, the HyperCard stack had a "status" box that displayed current program activity. For example, "Ready" indicated that the program was prepared to answer calls. A standard HyperCard "Home" button appeared at the top of the stack card. A computer icon button at the bottom of stack card, when clicked once, initiated a program restart action. Figure 1 illustrates the HyperCard stack.

The computer program administered the CES-D over the telephone by playing prerecorded digitized prompts and employing speaker-dependent speech recognition to create interactivity. To generate a simpler and more intuitive oral response to the CES-D items, we converted the response format from the standard four choices to eight choices (the actual number of days, including zero). In addition, we added the phrase "Again" to give the respondent the opportunity to repeat a CES-D item. Prior to answering the items, the program required a training segment to build a "template" of the respondent's speech characteristics for each brief discrete CES-D choice. During template training, the program prompted the respondent to repeat each discrete phrase three times. After obtaining the spoken input, the program averaged the three repetitions to create and store a template for each phrase. The training segment of the program lasted approximately 3–4 min. The program then proceeded to instructions for completing the CES-D. During administration of the CES-D, the program presented each item individually and waited for the participant's response. Subsequent to a spoken response, the program used the template to match and score a response. Alternatively, the participant could use the corresponding telephone Touch-Tone digit for a response that was automatically recorded by the program. Upon completion of the items, an interpretive report was printed at the university facility. A summary of the entire computer-telephone interview sequence is provided in the Appendix.

Design

The study employed a single-session counterbalanced 2 × 2 (language × order) experimental design. Random assignment of the three depression instruments controlled for order effects. The actual cell sizes, however, did not distribute equally. For 14 of the Spanish-speaking participants, the telephone method preceded the face-to-face screening; for 16, the face-to-face method came first. For 10 English-speaking participants, the telephone method preceded the face-to-face interview; for 12 the face-to-face interview came first. Participant acceptability included acceptance ratings of each CES-D method and preference rates between the two methods. Psychometric analyses of the telephone and face-to-face methods assessed, by language, the equivalence of total score means and variances as well as reliability estimates for alternate forms and internal consistency. The depression-symptom checklist served as an independent validity measure for the CES-D methods.

Table 1
Demographic Characteristics of the English and Spanish Language Groups

Characteristic	Language Group			
	English		Spanish	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age (years)	36.59	11.02	26.70	8.25
Education (years)	12.91	2.07	7.10	3.77
Computer experience*	2.32	1.13	1.48	0.80

*Rating: 1–5 (1 = *no experience*, 5 = *very knowledgeable*).

Figure 1. HyperCard stack of the computerized speech-recognition cellular telephone CES-D prototype.

RESULTS

Operational Issues

Of the total participant responses registered by the computer–telephone method, 95% were spoken. Largely as a result of less than perfect speech-recognition performance, similar proportions of the English (27%) and Spanish (23%) language groups reverted to Touch-Tone digits, at least once, in place of a spoken response. These limitations were associated with changes in respondent tone and pitch, which is known to reduce the accuracy of speaker-dependent speech-recognition systems (Noyes, Haigh, & Starr, 1989). In addition, the interviewer had to call back and repeat the computer–telephone program, at least once, for nearly half of the subjects (45% and 47% in the English and Spanish language groups, respectively), primarily because of inadequate template training. For example, a poor template error halted the program and required restarting the program for training. Communication disruptions, such as phone disconnections, accounted for less than 5% of the cases, although static occasionally diminished the compatibility of the computer–telephone screening.

Acceptability

After completing each CES-D method, every participant gave an acceptance rating of the method. Table 2 shows that both language groups positively rated the methods in-

dividually. After completing both methods, each participant was asked for a preference between the two. An analysis of the CES-D method preference revealed that 76% of the English speakers and 60% of the Spanish speakers preferred the computer–telephone method over the face-to-face one. A chi-square analysis revealed that the English sample's preference for the computer–telephone mode was significant [$\chi^2(1, N = 21) = 5.76, p < .02$]. Many English speakers (53%) reported preferring the computer–telephone mode because it seemed more personable. The Spanish speakers viewed the computer method as being more comfortable (33%) and easier to understand (28%).

Administration Time

Total administration time included computerized introduction, template training, CES-D instructions, item presentation, closing remarks, scoring, and the saving of responses. A correlated-samples *t* test for each language group indicated that the telephone administration time was significantly longer than the face-to-face method for both the English [$t(20) = 15.91, p < .01$, two-tailed] and the Spanish-speaking samples [$t(27) = 8.32, p < .01$, two-tailed]. The CES-D item administration times for the telephone method, which excluded the template training, were also analyzed. A correlated-samples *t* test indicated that there were no significant differences in the item-administration times between the methods for the English [$t(20) = 1.15$, two-tailed] and Spanish language groups [$t(27) = .13$, two-tailed]. The results suggested that when template training was excluded, the two methods were comparable in administration times (see Table 2).

Psychometric Properties

We analyzed the computer–telephone CES-D data by recoding the 0-to-7 days responses to match the standard CES-D categories. For example, a "7" days response was converted to "5 to 7 days" and scored as "3." A repeated measures multivariate analysis of variance (MANOVA) for language $\times$ order on the scores of the two CES-D methods found no order effects [$F(1,48) = 2.10$, n.s.]. However, significant main effects occurred between language [$F(1,48) = 17.37, p < .001$] and across methods [$F(1,48) = 49.86, p < .001$]. A dependent-sample *t* test for heterogeneity of variance between the computer–telephone and face-to-face scores indicated that the variances were not significantly different for the English [$t(21) = 1.73$, two-tailed] and Spanish [$t(29) = .15$, two-tailed] language

Table 2
Descriptive Statistics of CES-D Methods by Language Group

Variable	English		Spanish	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Computer–telephone acceptance ratings	6.55	2.30	7.33	1.71
Face-to-face acceptance ratings	7.46	1.50	8.00	1.73
Computer–telephone CES-D total time (minutes)	4.92	0.98	4.97	0.60
Face-to-face CES-D time (minutes)	3.17	1.32	2.85	1.17
CES-D item-only time (minutes)	2.80	0.93	2.85	0.60
Computer–telephone CES-D recoded total score	24.91	12.89	14.60	5.58
Face-to-face CES-D total score	21.32	14.78	8.60	5.68

Table 3
Correlations of the CES-D Methods by Language Group

Variable	English		Spanish	
	Phone	Face	Phone	Face
Internal consistency reliability (α)	.78	.82	.81	.72
Intercorrelation with depression checklist (r)	.70*	.73*	.59*	.60*
Alternate form reliability between both methods (r)	.92*		.76*	

* $p < .001$

groups. Although the variances were homogeneous for the two groups, both groups scored significantly higher on the computer–telephone method.³ Table 2 summarizes the means.

Interitem consistency analyses conducted on the CES–D methods did not involve recoding the responses, including the four nondepression items. The results yielded high α coefficients on the telephone method for both language samples. The face-to-face-method α estimates were high for the English speakers and moderately high for the Spanish speakers. The significantly high coefficients of equivalence between the computer–telephone and face-to-face CES–D scores for the English and Spanish speakers offered support for alternate-form reliability. The intercorrelations of the depression checklist total scores with English computer–telephone and face-to-face scores were similarly high. The difference between the coefficients was not statistically significant [$t(19) = .46$, two-tailed]. The Spanish computer–telephone and face-to-face correlations were moderately high, but the difference was not statistically significant [$t(27) = .08$, two-tailed]. Table 3 summarizes the coefficients. The similar correlations added evidence to the psychometric validity of the two CES–D methods.⁴

DISCUSSION

The results of our exploratory study suggested that the computer–telephone method adequately administered a depression scale for both language groups. The positive participant acceptance of the prototype was consistent with the literature on computerized assessment (Lukin, Dowd, Plake, & Kraft, 1985; Rozensky, Honor, Rasinski, Tovian, & Herz, 1986). In addition, participant preference for the prototype replicated previous research suggesting that some respondents attribute human qualities to an interactive computer program (Muñoz et al., 1991). Furthermore, the strong psychometric properties of the prototype supported its reliability and validity, which paralleled previous findings on computer-assisted assessment (Honaker et al., 1988; Kobak et al., 1990; Wilson et al., 1985).

The total-score mean differences between the CES–D modes, regardless of language, possibly reflected the “mean shift” suggested by Marco (1981, cited in Hofer & Green, 1985), that is, that adapted computerized test scores might be affected by a constant quantity. Thus, a constant may need to be added to or subtracted from the average. In our initial sample, the computer–telephone method displayed a positive shift from the face-to-face mean of about 4 points for the English speakers and 6 points for the Span-

ish speakers. The differences between the initial computer–telephone and face-to-face scores also raised questions concerning the relationship of participants’ levels of self-disclosure to the CES–D mode. Previous research had evidenced respondents’ underreporting of socially sensitive behavior in face-to-face interviewing (Catania, McDermott, & Pollack, 1986). While lower rates of self-disclosure among Latinos, as compared with non-Hispanic whites, occurred under oral interview conditions (LeVine & Franco, 1981), it is possible that respondents may be more guarded with a similar ethnicity interviewer because of the potential for future interactions (Franco, Malloy, & González, 1984). Since the depression checklist was initially administered face-to-face, however, the change in self-disclosure could not be adequately assessed.

In our follow-up study, described in note 4, there were no differences in total score means and variances between the methods, and the paper-and-pencil depression checklist correlated similarly with the CES–D methods. These results suggested that the differences were largely associated with the item-response formats, not with the actual modes of presentation. Standard response formats need to be adapted to make computerized assessment applications suitable for special populations (Carr, Wilson, Ghosh, Ancill, & Woods, 1982). Flexible and intuitive oral responses were required for the all-audio speech-recognition computer–telephone prototype. Therefore, special modifications compel the exploration of restandardizing norms and cutoff scores for a variety of computerized psychological assessment techniques (Hofer & Green, 1985), including speech-recognition telephone applications.

The samples in this exploratory study were small and chiefly self-selected from a few sampled field sites. Obviously, the results cannot be generalized without larger and more representative randomized samples. The lack of a retest group for assessing test–retest reliability also limited the statistical power of the analysis (Honaker, 1988). As previously noted, we had hypothesized that there would be no differences between CES–D methods. This placed our study, as with all investigations of equivalency between groups, in the precarious situation of confirming the null hypothesis. Rogers, Howard, and Vessey (1993) argue that equivalency testing between experimental groups also needs to involve a “nonequivalence” null hypothesis, and they propose an alternative hypothesis for “equivalence.” Our study also lacked a measure of acculturation for the Spanish speakers (Cuellar, Harris, & Jasso, 1980; Marín, Sabogal, Marín, Otero-Sabogal, & Pérez-Stable, 1989). The Spanish-speaking sample reported significantly less computer experience and fewer years of education, and

may have been less acculturated as well. Exploring the relationship between the levels of acculturation and computerized scores would enhance the data interpretation (Marín & Marín, 1991).

One potential advantage of computerized interviewing is that administration times may be comparable to or faster than conventional methods (White, Clements, & Fowler, 1985). Speaker-independent continuous speech-recognition technology based on language syntax and sound, does not require template training (Bergeron, 1991). Eliminating template training would make speech-recognition assessment more rapid and efficient by removing inadequate templates and reducing misrecognitions. Beyond computerized self-report instruments, future possibilities lie in the development and testing of objective computerized measures for screening depression using spectral analysis of voice samples. Depressed persons may demonstrate clinical markers of speech differences (Hargreaves & Starkweather, 1964; Vanger, Summerfield, Rosen, & Watson, 1992). Scherer and Zei (1988) found that lower pitch was associated with higher levels of depression. Previous research had also indicated that depressed individuals display differences in response latency times (Mandal, Srivastava, & Singh, 1990). Thus, computerized interactive-speech programs may provide clinical evidence of depression (Starkweather, 1992). Given the current findings and potential future developments, the computer cellular telephone method remains promising as an alternative for presenting a depression-screening measure.

REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- ANDERSEN, M. L. (1993). Studying across difference: Race, class, and gender in qualitative research. In J. H. Stanfield II & R. M. Dennis (Eds.), *Race and ethnicity in research methods* (pp. 39-52). Newbury Park, CA: Sage.
- ARTICULATE SYSTEMS (1993). *Voice navigator 2.3.2* [Computer program]. Woburn, MA: Author.
- BERGERON, B. (1991). Challenges associated with providing speech recognition user interfaces for computer-based educational systems. *Collegiate Microcomputer*, 4, 129-143.
- BROADHEAD, W. E., CLAPP-CHANNING, N. E., FINCH, J. N., & COPELAND, J. A. (1989). Effects of medical illness and somatic symptoms on treatment of depression in a family residency practice. *General Hospital Psychiatry*, 11, 194-200.
- CARR, A. C., GHOSH, A., & ANCILL, R. J. (1983). Can a computer take a psychiatric history? *Psychological Medicine*, 13, 151-158.
- CARR, A. C., WILSON, S. L., GHOSH, A., ANCILL, R. J., & WOODS, R. T. (1982). Automated testing of geriatric patients using a microcomputer-based system. *International Journal of Man-Machine Studies*, 17, 297-300.
- CATANIA, J. A., McDERMOTT, L. J., & POLLACK, L. M. (1986). Questionnaire response bias and face-to-face interview sample bias in sexuality research. *Journal of Sex Research*, 22, 52-72.
- CLARIS CORPORATION (1991). *HyperTalk 2.1* [Computer program]. Santa Clara, CA: Author.
- COMSTACK, G. W., & HELSING, K. J. (1976). Symptoms of depression in two communities. *Psychological Medicine*, 6, 551-563.
- CUELLAR, I., HARRIS, L. C., & JASSO, R. (1980). An acculturation scale for Mexican-American normal and clinical populations. *Hispanic Journal Behavioral Sciences*, 2, 199-217.
- CUNNINGHAM, J. K., ROBINSON, G. L., & SERPE, R. T. (1993). *Homeless persons in Orange county: Demographics, needs, and health-risk behaviors* (Document No. RDR-101). Santa Ana, CA: County of Orange Health Care Agency.
- FOWLER, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting & Clinical Psychology*, 53, 748-759.
- FRANCO, J. N., MALLOY, T., & GONZÁLEZ, R. (1984). Ethnic and acculturation differences in self-disclosure. *Journal of Social Psychology*, 122, 21-32.
- GONZÁLEZ, G. M. (1993a). Computerized speech recognition in psychological assessment: A Macintosh prototype for screening depressive symptoms. *Behavior Research Methods, Instruments, & Computers*, 25, 301-303.
- GONZÁLEZ, G. M. (1993b). A computerized speech recognition telephone application for screening clinical depression. In *Proceedings of the 17th Annual Symposium on Computer Applications in Medical Care* (p. 936). New York: McGraw-Hill.
- GONZÁLEZ, G. M., SPITERI, C. B., & KNOWLTON, J. (1995). A computerized speech recognition pilot study for screening depressive symptoms. *Computers in Human Behavior*, 11, 85-93.
- GREENBERG, P. E., STIGLIN, L. E., FINELSTEIN, S. N., & BERNDT, E. R. (1993). The economic burden of clinical depression in 1990. *Journal of Clinical Psychiatry*, 54, 405-418.
- HARGREAVES, W. A., & STARKWEATHER, J. A. (1964). Voice quality changes in depression. *Language Speech*, 7, 84-88.
- HOFER, P. J., & GREEN, B. F. (1985). The challenge of competence and creativity in computerized psychological assessment. *Journal of Consulting & Clinical Psychology*, 53, 826-838.
- HONAKER, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8, 561-577.
- HONAKER, L. M., HARRELL, T. H., & BUFFALO, J. D. (1988). Equivalency of microtest computer MMPI administration for standard and special scales. *Computers in Human Behavior*, 4, 323-337.
- HOUGH, R. L., LANDSVERK, J. A., KARNO, M., BURNAM, M. A., TIMBERS, D. M., ESCOBAR, J. I., & REGIER, D. A. (1987). Utilization of health and mental health services by Los Angeles Mexican-Americans and non-Hispanic whites. *Archives of General Psychiatry*, 44, 702-709.
- KOBAK, K. A., REYNOLDS, W. M., ROSENFELD, R., & GREIST, J. H. (1990). Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. *Psychological Assessment: A Journal of Consulting & Clinical Psychology*, 2, 56-63.
- LAVRAKAS, P. J. (1987). *Telephone survey methods: Sampling, selection, and supervision*. Newbury Park, CA: Sage.
- LEVINE, E., & FRANCO, J. N. (1981). A reassessment of self-disclosure patterns among Anglo-Americans and Hispanics. *Journal of Counseling Psychology*, 28, 522-524.
- LEVINE, S., ANCILL, R. J., & ROBERTS, A. P. (1989). Assessment of suicide risk by computer-delivered self-rating questionnaire: Preliminary findings. *Acta Psychiatrica Scandinavica*, 80, 216-220.
- LUKIN, M. E., DOWD, E. T., PLAKE, B. S., & KRAFT, R. G. (1985). Comparing computerized versus traditional psychological assessment. *Computers in Human Behavior*, 1, 49-58.
- MANDAL, M. K., SRIVASTAVA, P., & SINGH, S. K. (1990). Paralinguistic characteristics of speech in schizophrenics and depressives. *Journal of Psychiatric Research*, 74, 191-196.
- MARÍN, G., & MARÍN, B. V. (1989). A comparison of three interviewing techniques for studying sensitive topics with Hispanics. *Hispanic Journal of Behavioral Sciences*, 11, 330-340.
- MARÍN, G., & MARÍN, B. V. (1991). *Research with Hispanic populations*. Newbury Park, CA: Sage.
- MARÍN, G., PÉREZ-STABLE, E. J., & MARÍN, B. V. (1989). Cigarette smoking among San Francisco Hispanics: The role of acculturation and gender. *American Journal of Public Health*, 79, 196-198.
- MARÍN, G., SABOGAL, F., MARÍN, B. V., OTERO-SABOGAL, R., & PÉREZ-STABLE, E. J. (1989). Development of a short acculturation scale for Hispanics. *Hispanic Journal of Behavioral Sciences*, 9, 183-205.
- MOORE, N. C., SUMMER, K. R., & BLOOR, R. N. (1984). Do patients like psychometric testing by computer? *Journal of Clinical Psychiatry*, 40, 875-877.
- MOSCIKI, E. K., LOCKE, B. Z., RAE, D. S., & BOYD, J. H. (1989). Depressive symptoms among Mexican Americans: The Hispanic health

- and nutrition examination survey. *American Journal of Epidemiology*, **120**, 348-360.
- MUÑOZ, R. F. (1993). Depression prevention: Current research and practice. *Applied & Preventive Psychology*, **2**, 21-33.
- MUÑOZ, R. F., GONZÁLEZ, G. M., & STARKWEATHER, J. (1991, August). *Automated screening for depression using computerized speech recognition*. Paper presented at the meeting of the American Psychological Association, San Francisco.
- MUÑOZ, R. F., & YING, Y. (1993). *The prevention of depression: Research and practice*. Baltimore: Johns Hopkins University Press.
- NOYES, J. M., HAIGH, R., & STARR, A. F. (1989). Automatic speech recognition for disabled people. *Applied Ergonomics*, **20**, 293-298.
- PÉREZ-STABLE, E. J., MIRANDA, J., MUÑOZ, R. F., & YING, Y. W. (1990). Depression in medical outpatients: Underrecognition and misdiagnosis. *Archives of Internal Medicine*, **150**, 1083-1088.
- RADLOFF, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 385-401.
- RICHARDS, J. S., FINE, P. R., WILSON, T. L., & ROGERS, J. T. (1983). A voice-operated method for administering the MMPI. *Journal of Personality Assessment*, **47**, 167-170.
- ROBERTS, R. E. (1980). Reliability of the CES-D scale in different ethnic contexts. *Psychiatry Research*, **2**, 125-134.
- ROBINS, L. N., HELZER, J. E., ORVASCHEL, H., ANTHONY, J. C., BLAZER, D. G., BURNAM, A., & BURKE, J. D., JR. (1985). In W. W. Eaton & L. G. Kessler (Eds.), *Epidemiological field methods in psychiatry: NIMH Epidemiological Catchment Area program* (pp. 238-260). New York: Academic Press.
- ROGERS, J. L., HOWARD, K. I., & VESSEY, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, **113**, 553-565.
- ROZENSKY, R. H., HONOR, L. F., RASINSKI, K., TOVIAN, S. M., & HERZ, G. I. (1986). Paper-and-pencil versus computer-administered MMPIs: A comparison of patients' attitudes. *Computers in Human Behavior*, **2**, 111-116.
- SCHERER, K. R., & ZEI, B. (1988). Vocal indicators of affective disorders. *Psychotherapy & Psychosomatics*, **49**, 179-186.
- SHAPIRO, S., SKINNER, E., KESSLER, L., VON KORFF, M., GERMAN, P., TISCHLER, G., LEAF, P. J., BENHAM, L., COTTLER, L., & REGIER, D. A. (1984). Utilization of health and mental health services: Three Epidemiological Catchment Area sites. *Archives of General Psychiatry*, **41**, 971-978.
- STARKWEATHER, J. A. (1992). Computer applications in psychiatric interviewing. In K. C. Lun et al. (Eds.), *Proceedings of the MediInfo 92* (p. 318). Amsterdam: Elsevier, North-Holland.
- STARKWEATHER, J. A., & MUÑOZ, R. F. (1989, May). *Identification of clinical depression among foreign speakers*. Paper presented at the meeting of the American Association for Medical Systems and Informatics, San Francisco.
- STOUDEMIRE, A., FRANK, R., KAMLET, M., & HEDEMARK, N. (1987). Depression. In R. W. Amler & H. B. Dull (Eds.), *Closing the gap: The burden of unnecessary illness* (pp. 65-72). New York: Oxford University Press.
- VANGER, P., SUMMERFIELD, A. B., ROSEN, B. K., & WATSON, J. P. (1992). Effects of communication on speech behavior of depressives. *Comprehensive Psychiatry*, **33**, 39-41.
- WEISSMAN, M. M., BRUCE, M. L., LEAF, P. J., FLORIO, L. P., & HOLZER, C. (1991). Affective disorders. In L. N. Robins & D. A. Regier (Eds.), *Psychiatric disorders in America: The Epidemiological Catchment Area study* (pp. 53-79). New York: Free Press.
- WEISSMAN, M. M., SHOLOMSKAS, D., POTTINGER, M., PRUSOFF, B. A., & LOCKE, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology*, **106**, 203-214.
- WHITE, D. M., CLEMENTS, C. B., & FOWLER, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior*, **1**, 153-162.
- WILSON, F. R., GENCO, K. T., & YAGER, G. G. (1985). Assessing the equivalence of paper-and-pencil vs. computerized tests: Demonstration of a promising technology. *Computers in Human Behavior*, **1**, 265-275.

NOTES

1. "Equivalency" studies comparing computerized and conventional assessment methods propose confirmation of the null hypothesis, since significant differences suggest that the methods are not equivalent.
2. We noted that our limited sampling influenced these striking contrasts, and we therefore interpreted differences between the language groups with caution. Reporting the similarities or differences between the language groups was for the purpose of comparison and not for generalization.
3. Post hoc analyses of the responses to the standard face-to-face and adapted computer-telephone formats suggested that participants in both language groups more frequently endorsed "less than 1 day" with the face-to-face format [$F(1,50) = 9.25, p < .005$], but more frequently endorsed "1 or 2 days" with the computer-telephone format [$F(1,50) = 8.23, p < .01$]. This tended to inflate the computer-telephone scores.
4. To further clarify the analysis of the elevated telephone CES-D scores and to discern whether the results were related to the actual method or response format, we collected additional data. Eighteen English- and 10 Spanish-speaking university students completed randomly ordered equivalent computer-telephone and face-to-face CES-D methods that employed 0-to-7-day response formats. The research procedures were the same as before, except that the interviews were conducted at a confidential university setting and participants completed the depression checklist with paper and pencil. The latter modification was intended to assess change in self-disclosure between the face-to-face and telephone methods. A repeated measures MANOVA for language $\times$ order on the recoded CES-D total scores indicated that there were no significant main effects across the methods [$F(1,24) = 1.62$]. Furthermore, there were no differences in total score variances for the English [$t(16) = 0.55$, two-tailed] and Spanish-speaking [$t(8) = 0.90$, two-tailed] groups. Correlations were also computed for the combined English- and Spanish-speaking university-student sample. The coefficient of equivalence between the two CES-D methods was significant [$r(28) = .83, p < .001$]. The intercorrelations with the paper-and-pencil depression checklist were also significant for the computer-telephone [$r(28) = .60, p < .001$] and face-to-face [$r(28) = .71, p < .001$] methods, but the difference between the two coefficients was not statistically significant [$t(25) = 1.20$, two-tailed].

APPENDIX

Summary of the Computer-Telephone Interview Sequence

1. Interviewer
 - a. Presents respondent with oral instructions for completing computer-telephone method
 - b. Calls the computer, enters the respondent's identification number, and selects the language
 - c. Hands the cellular telephone to the respondent
2. Computer-Respondent Interaction
 - a. Computer presents an introduction and informs the respondent to train a voice template
 - b. Respondent trains a voice template
 - c. Computer builds and stores a respondent voice template
 - d. Computer presents instructions for completing the CES-D and instructs the respondent to verbally answer each item
 - e. Computer presents a CES-D item
 - f. Respondent verbally responds to each item
 - g. Upon completing the 20 CES-D items, the computer thanks the respondent, requests the interviewer be advised, and hangs up
3. Computer
 - a. Scores the responses and saves the results
 - b. Prints the results to a report form at the university facility

(Manuscript received July 5, 1994;

revision accepted for publication September 19, 1994.)