

The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect

STEFAN E. SCHULENBERG and BARBARA A. YUTRZENKA
University of South Dakota, Vermillion, South Dakota

The use of computerized psychological assessment is a growing practice among contemporary mental health professionals. Many popular and frequently used paper-and-pencil instruments have been adapted into computerized versions. Although equivalence for many instruments has been evaluated and supported, this issue is far from resolved. This literature review deals with recent research findings that suggest that computer aversion negatively impacts computerized assessment, particularly as it relates to measures of negative affect. There is a dearth of equivalence studies that take into account computer aversion's potential impact on the measurement of negative affect. Recommendations are offered for future research in this area.

With the continuous, rapid advent of computerized assessment products for the contemporary mental health professional, mental health disciplines are in the midst of a technological crossroads. Computer technology has proven to be a major force in the practice of psychiatry and psychology (Johnson, 1984; Kobak, Reynolds, & Greist, 1994), as well as in the related fields of social work (Nurius & Hudson, 1989; Pardeck & Schulte, 1990) and rehabilitation (Burkhead & Sampson, 1985). Specialized subfields of psychology that have been impacted by computer technology include forensic psychology (Gummow, 1991), clinical psychology (Brown, 1984), school psychology (Brown, 1984; Kramer & Gutkin, 1990), and neuropsychology (Golden, 1987).

It has long been thought that computers may benefit the psychological assessment of individuals (Smith, 1963). Some have described the advancement of computer technology as unstoppable (Lukin, Dowd, Plake, & Kraft, 1985) and irreversible (Merton & Ruch, 1996), and the application of computers to psychological testing has been indicated to be extensive (Lanyon, 1987) and widespread (Johnson, 1984). In a survey of 227 psychologists, Farrell (1989) found that 55% used a computer in their practice, with an additional 35% planning on utilizing computers within a year. Farrell noted that 44% of the 130 who use computers in their practice did so at least occasionally for the purpose of test administration. Computers are highly attractive to psychologists because of their decreasing cost, increasing power, and increasing availability (Sturges, 1998; Tseng, Macleod, & Wright, 1997). It is likely that

their use in computerized psychological assessment will continue to become more pervasive as time passes (Ford, Vitelli, & Stuckless, 1996).

Computers have been integrated into psychological assessment to the extent that many popular and frequently used paper-and-pencil instruments have been adapted into a computerized version. Equivalence for many instruments has been evaluated, but it seems clear that this issue is far from being resolved. Although many have written of the pros and cons of computerized versus conventional assessment (see Space, 1981, for a review), this paper will not address these concerns. Rather, the purpose of this paper is to address a more recent finding in the literature that suggests a link between computer aversion and the measurement of negative affect. That is, the literature suggests that those who experience discomfort when using computers may have higher scores on computerized measures of negative affect than those who do not experience such discomfort (see George, Lankford, & Wilson, 1992; Peterson, Johannsson, & Carlsson, 1996). The literature review for this article was conducted using a variety of search techniques, including searching a computer database (PSYCLIT) using various combinations of search terms (e.g., computer assessment, computer aversion, equivalence, paper-and-pencil, etc.) and perusing reference pages of articles. Attempts were also made to contact some of the researchers in the field, inquiring as to additional research that had been conducted, or was currently being conducted, relating to computer aversion and/or instruments of negative affect.

Defining Applicable Terminology

Equivalence. When one is referring to equivalence between paper-and-pencil and computerized versions of a test, one is essentially asking whether the instruments represent alternate forms of the test in question (Harrell, Honaker, Hetu, & Oberwager, 1987). Unanticipated con-

This review is a portion of the first author's master's thesis research. The authors acknowledge the contributions of three anonymous reviewers with regard to editing drafts of this manuscript. Correspondence concerning this article should be addressed to S. Schulenberg, Department of Psychology, University of South Dakota, 414 E. Clark Street, Vermillion, SD 57069 (e-mail: sschulen@usd.edu).

sequences may arise, no matter how fine the adaptation (Kubinger, Formann, & Farkas, 1991). Furthermore, the process involved in determining whether error may have been introduced as a consequence of altering the method of test administration is highly complex (Brown, 1984). Until the equivalence of a particular administration format is demonstrated empirically, the validity of a computerized instrument remains unknown (Ford et al., 1996). Furthermore, if equivalence between computerized and conventional instruments has not been empirically supported, psychometric properties, such as norms, reliability, and validity, cannot be adapted from conventional measures to aid in the interpretation of computerized instruments (Tseng et al., 1997).

It is imperative to establish a firm empirical base, to ensure the equivalency of instruments across modalities. According to Lukin et al. (1985) and Moreland (1985), equivalency studies of this sort exist in the literature but are relatively few in number. However, these authors differ in their impression of the results of these studies. Lukin et al. reported that the correlational information available appears to be supportive of equivalence across modalities. In contrast, Moreland described the literature in this area as being inconsistent. Moreland reached a similar, albeit provisional, conclusion that nonequivalence between computerized and paper-and-pencil measures is statistically insignificant, if it is to be found at all. Others have also noted the ambiguous information available on equivalence (e.g., Burke & Normand, 1987; Lockshin & Harrison, 1991). One reason for the ambiguity in the literature may be different definitions as to what constitutes equivalence.

According to the American Psychological Association, in its *Guidelines for Computer-Based Tests and Interpretations* (American Psychological Association, 1986), equivalence between paper-and-pencil and computerized tests may be determined if "(a) the rank orders of scores of individuals tested in alternative modes closely approximate each other, and (b) the means, dispersions, and shapes of the score distributions are approximately the same, or have been made approximately the same by re-scaling the scores from the computer mode" (p. 18). Some have advocated the development of fresh norms for psychological tests if nonequivalence has been determined (American Psychological Association, 1986; Kubinger et al. 1991; Pineseault, 1996; Vansickle & Kapes, 1993), whereas others have suggested using an equating formula (Hofer, 1985). Given that mean differences in alternate modes of the same instrument may be accounted for by establishing different norms, the authors subscribe to a different definition that may be considered as defining equivalence—equivalence in construct validity. Computerization may change a psychological test to the extent that computerized and written versions of the same test may not be measuring the same construct (Moreland, 1985). For example, for the purpose of this paper, the question of equivalence relates to whether a

computerized adaptation of a paper-and-pencil depression instrument is actually measuring depression or whether computerization somehow detracts from the measurement of depression, because of computer aversion. Others have noted the viability and importance of construct validity (e.g., King & Miles, 1995; Neuman & Baydoun, 1998; Turban, Sanders, Francis, & Osburn, 1989).

Although alternative definitions of equivalence may result in ambiguity, other explanations have been offered for the mixed results of equivalency studies. George et al. (1992) noted that variability may be due to the fact that many studies are characterized by weaknesses in methodology (e.g., small sample sizes, low power to detect mean disparities). These authors indicated that the failure to account for individual differences (e.g., computer anxiety) may be a significant factor in understanding mean differences.

Computer aversion. George et al. (1992) noted that, at the time of their study, no research existed that systematically evaluated the impact of computer anxiety on computerized psychological assessment. Gardner, Disenza, and Dukes (1993) described computer anxiety as being a major cause of resistance to using computers, and Dimock and Cormier (1991) argued that individual differences in computer anxiety across modalities may yield stark disparities in the results.

Meier and Lambert (1991) noted that people who demonstrate discomfort with computers have been described in the literature as having an *aversion* toward computers, a *phobia* of computers, or an *anxiety* toward computers. Meier and Lambert argued against the use of such terms as *phobia* and *anxiety*, because the discomfort with computers should not be confused with severe psychological conditions. They emphasized the use of the term *computer aversion*, which is the term of choice when describing this phenomenon in this paper. Many instruments are available for the assessment of computer aversion. LaLomia and Sidowski (1993) reviewed the literature and provided an in-depth analysis of 12 such instruments (every computer aversion scale available at the time of their study). Many of the scales were noted to lack a sound theoretical foundation. In fact, the authors determined that, of the instruments examined, only the Computer Aversion Scale (CAVS; Meier, 1985, 1988) appeared to be grounded in social learning theory.

Computerized Assessment Instruments:

General Overview

Many commonly used paper-and-pencil instruments have been adapted for computerized administration. Research relating to such adaptations have included personality inventories (Pineseault, 1996; Rasulis, Schulberg, & Murtagh, 1996; Watson, Thomas, & Anderson, 1992), intelligence measures (Elwood & Clark, 1978; Katz & Dalby, 1981; Kubinger et al., 1991), problem checklists and scales of self-concept (Hinkle, Sampson, & Radonsky, 1991; Simola & Holden, 1992), career and

aptitude instruments (Harrell et al., 1987; Reardon & Loughead, 1988; Vansickle & Kapes, 1993), and neuropsychological instruments (Berger, Chibnall, & Gfeller, 1994, 1997; Choca & Morris, 1992), to name a few. It should again be noted that the definition as to what constitutes equivalence varies from study to study. However, the latter studies suggest a trend that largely supports the equivalence of conventional and computerized instruments. Two notable exceptions are Watson et al. (1992) and Kubinger et al. (1991). Watson et al., conducting meta-analyses of studies comparing conventional and computerized versions of the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1940, 1943), noted the potential for computerization to underestimate scores. Kubinger et al. found that the German version of Raven's Standard Progressive Matrices (Kratzmeier & Horn, 1987) was associated with lower IQ scores (13 IQ points, on average).

Although there are many equivalence studies across a wide variety of areas available in the literature to those interested, one area that is growing relates to the equivalence of instruments of negative affect. Table 1 presents a summary of the equivalence studies that relate to instruments of negative affect. The results of these studies suggest strong support for the use of computerized adaptations of paper-and-pencil instruments; however, clinicians must be wary, because the issue is more complex than it may initially appear. Nurius (1990) noted that there is a lack of research concerning response disparities owing to anxiety and/or unfamiliarity with computers. Brown (1984) indicated that individuals suffering from depressive symptomatology may have trouble dealing with a computer. Others have noted that computer anxiety may negatively affect a computerized assessment (Ford et al., 1996; George et al., 1992). Peterson et al. (1996) found in their research that, although there were no significant differences between the computer and the paper-and-pencil modalities, computerized Beck Depression Inventory (BDI) scores tended to be higher. Similar trends were found by Glaze and Cox (1991) and Schulenberg (1998), although these findings were also not statistically significant. Peterson et al. indicated that their finding warranted future study. They posed the question as to whether the computerized version solicited more honest information, owing to the lessened impact of social desirability. Peterson et al. also noted, however, that it is possible that increased computer scores may be a consequence of the computerized administration itself. Peterson et al. further indicated that any disparities between paper-and-pencil and computerized tests are cause for deliberation when normative information is adapted from paper-and-pencil to computerized tests for the purpose of interpretation—at a minimum, for instruments asking about sensitive information.

Dimock and Cormier (1991) noted that individuals unfamiliar with computers will demonstrate greater anxiety when in proximity to a computer than will those who

already have a familiarity with computers. The authors stated that an increase in anxiety could impair performance during a computerized assessment. However, in their experiments, they found no evidence to indicate that subjects' level of computer experience or computer anxiety influenced the results. Dimock and Cormier concluded that the cause of format differences is elusive. Interestingly, these authors proposed that the novelty of the task may play a role in the disparity of their findings, because the difference disappeared when the subjects involved in the study took the test for the second time. They concluded that it was the repeated testing that alleviated differences in format, not their process of computer familiarization.

George et al. (1992) conducted a study with the paper-and-pencil and computerized counterparts of the BDI and the State-Trait Anxiety Inventory (STAI) with 97 undergraduate students. The Computer Anxiety Rating Scale (CARS) was used to gather ratings of computer anxiety. The authors found statistically significant disparities in terms of means on the BDI and the State scale of the STAI. Their results led them to conclude that the two formats are statistically not equivalent and that there is variability (e.g., computer anxiety) of which individuals may be unaware. Importantly, they found that as subjects' computer aversion increased, depression scores increased in direct proportion. This was only the case in the computerized condition. The authors also noted that results obtained via computerized assessment may be difficult to analyze, as an individual's aversion toward the computer may cloud any attempts at interpretation. They have also indicated that the extrapolation of test scores from paper-and-pencil measures to computer measures is not an acceptable course. To combat this phenomenon, George et al. argued that separate norms need to be developed for computer versions of conventionally administered instruments, and computer aversion ratings need to be assessed prior to the interpretation of any results garnered from a computer administration. Similarly, Tseng et al. (1997) argued that the computerized scores obtained in their sample were correlated with computer aversion, the end result being that elevations in mood are raised through computer aversion.

Implications for Measures of Negative Affect

As computerized assessment continues to become more pervasive, the available literature strongly suggests that the role of computer aversion, as it relates to computerized scores of negative affect, needs to be examined in greater detail. Instruments measuring negative affect need to be compared systematically with their computerized counterparts, to determine the equivalence of their respective construct validities. It should never be assumed that computerized psychological instruments are the same as their conventional counterparts, especially instruments of negative affect. Research must be conducted to determine and examine any differences that may exist across modalities.

Table 1
Equivalence Studies With a Focus on Instruments of Negative Affect

Study	Sample	Primary instruments	Findings
George, Lankford, and Wilson (1992)	97 college students	CARS (Heinssen, Glass, & Knight, 1987) paper-and-pencil only; BDI (Beck, 1967b); STAI (Spielberger, Gorsuch, & Lushene, 1970)	A MANOVA and univariate <i>t</i> tests revealed that BDI scores and State Anxiety scores were significantly higher in the computerized condition. No significant mean disparities were found for the Trait scale of the STAI.
Glaze and Cox (1991)	29 women from two community facilities	EPDS (Cox, Holden, & Sagovsky, 1987)	A correlation of .98 was found across modalities.
Kobak, Reynolds, and Greist (1993)	292 adults (78 controls and 214 outpatients)	HAS (Hamilton, 1959, 1967); CHAS (Kobak, Reynolds, & Greist, 1990)	Both forms had a coefficient alpha of .92, test-retest reliabilities of .96, and were strongly correlated (.92). Small but statistically significant disparities were noted with regard to mean scores and variances across the two forms.
Lukin, Dowd, Plake, and Kraft (1985)	66 college students	TRS (Dowd, Milne, & Wise, 1984); STAI; BDI (Beck, 1967a)	Tentative support was offered for equivalence of intake batteries across modalities.
Merten and Ruch (1996)	72 German-speaking adults were in the experimental group. They received a half of each questionnaire in a computer format, with the other half in a paper-and-pencil format.	EPQ-R (Eysenck & Eysenck, 1991); CRS (Carroll, Feinberg, Smouse, Rawson, & Greden, 1981)	Relating to the CRS, the findings suggested equivalence, in that no systematic disparities were found with regard to means and standard deviations for any of the groups.
Ogles, France, Lunnen, Bell, and Goldfarb (1998)	113 people (<i>M</i> age = 41.6 years)	BDI (Beck, 1978); BSD (Hakstian & McLean, 1989); CES-D (Radloff, 1977); CCES-D	The coefficient alpha for the CCES-D (.91) is comparable to the CES-D (.85 and .90 for nonclinical and clinical samples, respectively). The CCES-D and the CES-D are highly associated ($r = .96$). Concurrent validity of CCES-D with the BDI and the BSD are also significant ($r = .89$ and $r = .43$, respectively).
Peterson, Johannsson, and Carlsson (1996)	57 Swedish hospital patients	BDI (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961); MACL (Sjöberg, Svensson, & Persson, 1979); SPIQ (Rydberg & Höghliem, 1974),	Mean scores were determined to be comparable across modalities. A tendency for BDI scores to be higher with the computer administration was noted.
Rosenfeld, Dar, Anderson, Kobak, and Greist (1992)	70 adults; 23 non-patients, 31 meeting criteria for OCD, and 16 diagnosed with other anxiety disorders	Y-BOCS (Goodman et al., 1989a, 1989b)	Based on large correlations and means, the computer version is comparable with the conventional version with regard to measuring OCD symptomatology. The nonpatient sample and the sample with other anxiety disorders were more likely to describe more symptoms to the computer than were noted by clinicians.
Schulenberg (1998)	180 college students	BDI-II (Beck, Steer, & Brown, 1996); C-BDI-II; PMAQ (Merten & Ruch, 1996); CAVS	An independent samples <i>t</i> test analysis of mean scores suggested support for measurement equivalence.
Steer, Rissmiller, Ranieri, and Beck (1994)	330 inpatients diagnosed with various disorders	BDI (Beck & Steer, 1987), computer format; BHS (Beck & Steer, 1988), computer format	Findings suggested reliabilities and validities comparable with their paper-and-pencil counterparts.
Tseng, Macleod, and Wright (1997)	108 college students	VAMS (McCormack, de L. Horne, & Sheather, 1988), computer and conventional versions; CARS; Modified VMIP (Seibert & Ellis, 1991; Velton, 1968)	Computer anxiety scores correlated significantly with mood scores only when measured by the computer.

Note—BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory-II; BHS, Beck Hopelessness Scale; BSD, Brief Screen for Depression; CARS, Computer Anxiety Rating Scale; CAVS, Computer Aversion Scale; CES-D, Center for Epidemiological Studies-Depressed Mood Scale; CCES-D, Computerized Center for Epidemiological Studies-Depressed Mood Scale; CHAS, Computer Hamilton Anxiety Scale; CRS, Carroll Rating Scale for Depression; EPDS, Edinburgh Postnatal Depression Scale; EPQ-R, Eysenck Personality Questionnaire-Revised; HAS, Hamilton Anxiety Scale; MACL, Mood Adjective Checklist; PMAQ, Preference for Mode of Administration Questionnaire; SPIQ, a Swedish measure of intelligence; STAI, State-Trait Anxiety Inventory; TRS, Therapeutic Reactance Scale; VAMS, Visual Analog Mood Scales; VMIP, Velton Mood Induction Procedure; Y-BOCS, Yale-Brown Obsessive-Compulsive Scale.

Concerning the direction of future research, if computer aversion potentially raises problems, as the literature suggests, it seems clear that more must be learned about the concept of computer aversion. Consideration must be given to the samples used in this research, as well as the research designs used. For example, many studies use college students as subjects, yet today's college students are becoming more and more familiar with computers. This was evident in Schulenberg's (1998) recent investigation of the written and computerized versions of the BDI-II. The author attempted to systematically examine the impact of computer aversion on written and computerized BDI-II scores in a college student sample. A moderate negative correlation ($-.67$) was found between computer experience and computer aversion, and overall, the sample was computer experienced and tended not to experience aversion toward computers. People more likely to lack or to have limited computer experience must be targeted as research subjects, because they may be at greater risk of experiencing computer aversion. For example, Laguna and Babcock (1997) noted that older individuals tend not to have as much computer experience. They found that the older individuals in their sample described greater computer aversion than did their younger counterparts. Although they were interested in computerized cognitive tasks, their findings need to be clarified further, so that the effect on computerized measures of negative affect is demonstrated. Tseng et al. (1997) and Schulenberg found gender differences in their samples in descriptions of computer aversion, with women reporting higher levels. Gender differences such as these also need to be explored further. Additional research must also be conducted to see how computer aversion impacts clinical populations.

A second area of research that should be examined with regard to this phenomenon is research design. A variety of research designs have been used, and consistent with Steer, Rissmiller, Ranieri, and Beck's (1994) recommendation, counterbalanced designs are an effective way to go about research in this area. The use of counterbalanced experimental designs whereby individuals are given both versions of an instrument of negative affect—coupled with the accounting of variables such as computer experience, preference for method of administration (paper-and-pencil vs. computer), and computer aversion—may be most useful with regard to maximizing the detection of differences. Focusing research in these areas will likely maximize the quality of computerized psychological assessment services that continue to become more widely available to the larger society.

REFERENCES

- AMERICAN PSYCHOLOGICAL ASSOCIATION (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- BECK, A. T. (1967a). *Depression: Causes and treatments*. Philadelphia: University of Pennsylvania Press.
- BECK, A. T. (1967b). *Depression: Clinical, experimental and theoretical aspects*. New York: Harper & Row.
- BECK, A. T. (1978). *Depression inventory*. Philadelphia: Center for Cognitive Therapy.
- BECK, A. T., & STEER, R. A. (1987). *Manual for the revised Beck Depression Inventory*. San Antonio, TX: Psychological Corporation.
- BECK, A. T., & STEER, R. A. (1988). *Manual for the Beck Hopelessness Scale*. San Antonio, TX: Psychological Corporation.
- BECK, A. T., STEER, R. A., & BROWN, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- BECK, A. T., WARD, C. H., MENDELSON, M., MOCK, J., & ERBAUGH, J. (1961). An inventory for measuring depression. *Archives for General Psychiatry*, *4*, 561-571.
- BERGER, S. G., CHIBNALL, J. T., & GFELLER, J. D. (1994). The Category Test: A comparison of computerized and standard versions. *Assessment*, *1*, 255-258.
- BERGER, S. G., CHIBNALL, J. T., & GFELLER, J. D. (1997). Construct validity of the computerized version of the Category Test. *Journal of Clinical Psychology*, *53*, 723-726.
- BROWN, D. T. (1984). Automated assessment systems in school and clinical psychology: Present status and future directions. *School Psychology Review*, *13*, 455-460.
- BURKE, M. J., & NORMAND, J. (1987). Computerized psychological testing: Overview and critique. *Professional Psychology: Research & Practice*, *18*, 42-51.
- BURKHEAD, E. J., & SAMPSON, J. P. (1985). Computer-assisted assessment in support of the rehabilitation process. *Rehabilitation Counseling Bulletin*, *28*, 262-274.
- CARROLL, B. J., FEINBERG, M., SMOUSE, P. E., RAWSON, S. G., & GREDEN, J. F. (1981). The Carroll Rating Scale for Depression: I. Development, reliability and validation. *British Journal of Psychiatry*, *138*, 194-200.
- CHOCA, J., & MORRIS, J. (1992). Administering the Category Test by computer: Equivalence of results. *Clinical Neuropsychologist*, *6*, 9-15.
- COX, J. L., HOLDEN, J. M., & SAGOVSKY, R. (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*, 782-786.
- DIMOCK, P. H., & CORMIER, P. (1991). The effects of format differences and computer experience on performance and anxiety on a computer-administered test. *Measurement & Evaluation in Counseling & Development*, *24*, 119-126.
- DOWD, E. T., MILNE, C. R., & WISE, S. L. (1984, August). *The Therapeutic Reactance Scale: Development and reliability*. Paper presented at the annual convention of the American Psychological Association, Toronto.
- ELWOOD, D. L., & CLARK, C. L. (1978). Computer administration of the Peabody Picture Vocabulary Test to young children. *Behavior Research Methods & Instrumentation*, *10*, 43-46.
- EYSENCK, H. J., & EYSENCK, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult)*. London: Hodder & Stoughton.
- FARRELL, A. D. (1989). Impact of computers on professional practice: A survey of current practices and attitudes. *Professional Psychology: Research & Practice*, *20*, 172-178.
- FORD, B. D., VITELLI, R., & STUCKLESS, N. (1996). The effects of computer versus paper-and-pencil administration on measures of anger and revenge with an inmate population. *Computers in Human Behavior*, *12*, 159-166.
- GARDNER, D. G., DISCENZA, R., & DUKES, R. L. (1993). The measurement of computer attitudes: An empirical comparison of available scales. *Journal of Educational Computing Research*, *9*, 487-507.
- GEORGE, C. E., LANKFORD, J. S., & WILSON, S. E. (1992). The effects of computerized versus paper-and-pencil administration on measures of negative affect. *Computers in Human Behavior*, *8*, 203-209.
- GLAZE, R., & COX, J. L. (1991). Validation of a computerised version of the 10-item (self-rating) Edinburgh Postnatal Depression Scale. *Journal of Affective Disorders*, *22*, 73-77.
- GOLDEN, C. J. (1987). Computers in neuropsychology. In J. N. Butcher

- (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 344-354). New York: Basic Books.
- GOODMAN, W. K., PRICE, L. H., RASMUSSEN, S. A., MAZURE, C., FLEISCHMANN, R. L., HILL, C. L., HENINGER, G. R., & CHARNEY, D. S. (1989a). The Yale-Brown Obsessive Compulsive Scale: I. Development, use and reliability. *Archives of General Psychiatry*, **46**, 1006-1011.
- GOODMAN, W. K., PRICE, L. H., RASMUSSEN, S. A., MAZURE, C., FLEISCHMANN, R. L., HILL, C. L., HENINGER, G. R., & CHARNEY, D. S. (1989b). The Yale-Brown Obsessive Compulsive Scale: II. Validity. *Archives of General Psychiatry*, **46**, 1006-1011.
- GUMMOW, L. J. (1991). Use of the computer in forensic psychology. *American Journal of Forensic Psychology*, **9**, 5-31.
- HAKSTIAN, R. A., & MCLEAN, P. D. (1989). Brief screen for depression. *Psychological Assessment*, **1**, 139-141.
- HAMILTON, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, **32**, 50-55.
- HAMILTON, M. (1967). Diagnosis and rating of anxiety. *British Journal of Psychiatry*, **3**, 76-79.
- HARRELL, T. H., HONAKER, L. M., HETU, M., & OBERWAGER, J. (1987). Computerized versus traditional administration of the Multidimensional Aptitude Battery-Verbal scale: An examination of reliability and validity. *Computers in Human Behavior*, **3**, 129-137.
- HATHAWAY, S. R., & MCKINLEY, J. C. (1940). A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, **10**, 249-254.
- HATHAWAY, S. R., & MCKINLEY, J. C. (1943). *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation.
- HEINSEN, R. K., GLASS, C. R., & KNIGHT, L. A. (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior*, **3**, 49-59.
- HINKLE, J. S., SAMPSON, J. P., & RADONSKY, V. (1991). Computer-assisted versus paper-and-pencil assessment of personal problems in a clinical population. *Computers in Human Behavior*, **7**, 237-242.
- HOFER, P. J. (1985). Developing standards for computerized psychological testing. *Computers in Human Behavior*, **1**, 301-315.
- JOHNSON, J. H. (1984). An overview of computerized testing. In M. D. Schwartz (Ed.), *Using computers in clinical practice: Psychotherapy and mental health applications* (pp. 131-133). New York: Haworth Press.
- KATZ, L., & DALBY, J. T. (1981). Computer and manual administration of the Eysenck Personality Inventory. *Journal of Clinical Psychology*, **37**, 586-588.
- KING, W. C., & MILES, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology*, **80**, 643-651.
- KOBAK, K. A., REYNOLDS, W. M., & GREIST, J. H. (1990). *Hamilton Anxiety Rating Scale: Computer-administered version*. Unpublished computer interview.
- KOBAK, K. A., REYNOLDS, W. M., & GREIST, J. H. (1993). Development and validation of a computer-administered version of the Hamilton Anxiety Scale. *Psychological Assessment*, **5**, 487-492.
- KOBAK, K. A., REYNOLDS, W. M., & GREIST, J. H. (1994). Computerized and clinician assessment of depression and anxiety: Respondent evaluation and satisfaction. *Journal of Personality Assessment*, **63**, 173-180.
- KRAMER, J. J., & GUTKIN, T. B. (1990). School psychology, assessment, and the computer: An analysis of current relationships and future potential. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. VII, pp. 131-150). Hillsdale, NJ: Erlbaum.
- KRATZMEIER, H., & HORN, R. (1987). *SPM. RAVEN-MATRIZEN-TEST. Standard Progressive Matrices-Manual*. Weinheim: Beltz.
- KUBINGER, K. D., FORMANN, A. K., & FARKAS, M. G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices, in particular for computerized testing. *European Review of Applied Psychology*, **41**, 295-300.
- LAGUNA, K., & BABCOCK, R. L. (1997). Computer anxiety in young and older adults: Implications for human-computer interactions in older populations. *Computers in Human Behavior*, **13**, 317-326.
- LA LOMIA, M. J., & SIDOWSKI, J. B. (1993). Measurements of computer anxiety: A review. *International Journal of Human-Computer Interaction*, **5**, 239-266.
- LANYON, R. I. (1987). The validity of computer-based personality assessment products: Recommendations for the future. *Computers in Human Behavior*, **3**, 225-238.
- LOCKSHIN, S. B., & HARRISON, K. (1991). Computer-assisted assessment of psychological problems. In A. Ager & S. Bendall (Eds.), *Microcomputers and clinical psychology: Issues, applications and future developments* (pp. 47-63). Chichester, U.K.: Wiley.
- 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.
- MCCORMACK, H. M., DE L. HORNE, D. J., & SHEATHER, S. (1988). Clinical applications of visual analog scales: A critical review. *Psychological Medicine*, **18**, 1007-1019.
- MEIER, S. T. (1985). Computer aversion. *Computers in Human Behavior*, **1**, 171-179.
- MEIER, S. T. (1988). Predicting individual differences in performance on computer-administered tests and tasks: Development of the Computer Aversion Scale. *Computers in Human Behavior*, **4**, 175-187.
- MEIER, S. T., & LAMBERT, M. E. (1991). Psychometric properties and correlates of three computer aversion scales. *Behavior Research Methods, Instruments, & Computers*, **23**, 9-15.
- MERTEN, T., & RUCH, W. (1996). A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Personality & Individual Differences*, **20**, 281-291.
- MORELAND, K. L. (1985). Computer-assisted psychological assessment in 1986: A practical guide. *Computers in Human Behavior*, **1**, 221-233.
- NEUMAN, G., & BAYDOUN, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement*, **22**, 71-83.
- NURIUS, P. S. (1990). A review of automated assessment. *Computers in Human Services*, **6**, 265-281.
- NURIUS, P. S., & HUDSON, W. W. (1989). Computers and social diagnosis: The client's perspective. *Computers in Human Services*, **5**, 21-35.
- OGLLES, B. M., FRANCE, C. R., LUNNEN, K. M., BELL, M. T., & GOLDFARB, M. (1998). Computerized depression screening and awareness. *Community Mental Health Journal*, **34**, 27-38.
- PARDECK, J. T., & SCHULTE, R. S. (1990). Computers in social intervention: Implications for professional social work practice and education. *Family Therapy*, **17**, 109-121.
- PETERSON, L., JOHANNSSON, V., & CARLSSON, S. G. (1996). Computerized testing in a hospital setting: Psychometric and psychological effects. *Computers in Human Behavior*, **12**, 339-350.
- PINSONEAULT, T. B. (1996). Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. *Computers in Human Behavior*, **12**, 291-300.
- 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.
- RASULIS, R., SCHULDBERG, D., & MURTAGH, M. (1996). Computer-administered testing with the Rotter Incomplete Sentences Blank. *Computers in Human Behavior*, **12**, 497-513.
- REARDON, R., & LOUGHEAD, T. (1988). A comparison of paper-and-pencil and computer versions of the Self-Directed Search. *Journal of Counseling & Development*, **67**, 249-252.
- ROSENFELD, R., DAR, R., ANDERSON, D., KOBAK, K., & GREIST, J. H. (1992). A computer-administered version of the Yale-Brown Obsessive-Compulsive Scale. *Psychological Assessment*, **4**, 329-332.
- RYDBERG, S., & HÖGHIELM, R. (1974). *SPIQ: Snabbt performancetest på intelligens [SPIQ: A fast performance-test for intelligence]* [Manual]. Stockholm: Skandinaviska Testförlaget AB.
- SCHULENBERG, S. E. (1998). A comparison between the computerized and paper-and-pencil versions of the Beck Depression Inventory-II. Unpublished master's thesis, University of South Dakota, Vermillion.
- SEIBERT, P. S., & ELLIS, H. C. (1991). A convenient self-referencing mood induction procedure. *Bulletin of the Psychonomic Society*, **29**, 121-124.
- SIMOLA, S. K., & HOLDEN, R. R. (1992). Equivalence of computerized

- and standard administration of the Piers–Harris children's self-concept scale. *Journal of Personality Assessment*, **58**, 287-294.
- SJÖBERG, L., SVENSSON, E., & PERSSON, L.-O. (1979). The measurement of mood. *Scandinavian Journal of Psychology*, **20**, 1-18.
- SMITH, R. E. (1963). Examination by computer. *Behavioral Science*, **8**, 76-79.
- SPACE, L. G. (1981). The computer as psychometrician. *Behavior Research Methods & Instrumentation*, **13**, 595-606.
- SPIELBERGER, C. D., GORSUCH, R. L., & LUSHENE, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- STEER, R. A., RISSMILLER, D. J., RANIERI, W. F., & BECK, A. T. (1994). Use of the computer-administered Beck Depression Inventory and Hopelessness Scale with psychiatric inpatients. *Computers in Human Behavior*, **10**, 223-229.
- STURGES, J. W. (1998). Practical use of technology in professional practice. *Professional Psychology: Research & Practice*, **29**, 183-188.
- TSENG, H.-M., MACLEOD, H. A., & WRIGHT, P. (1997). Computer anxiety and measurement of mood change. *Computers in Human Behavior*, **13**, 305-316.
- TURBAN, D. B., SANDERS, P. A., FRANCIS, D. J., & OSBURN, H. G. (1989). Construct equivalence as an approach to replacing validated cognitive ability selection tests. *Journal of Applied Psychology*, **74**, 62-71.
- VANSICKLE, T. R., & KAPES, J. T. (1993). Comparing paper-pencil and computer-based versions of the Strong–Campbell Interest Inventory. *Computers in Human Behavior*, **9**, 441-449.
- VELTON, E. J. (1968). A laboratory task for induction of mood states. *Behavior Research & Therapy*, **6**, 473-482.
- WATSON, C. G., THOMAS, D., & ANDERSON, P. E. D. (1992). Do computer-administered Minnesota Multiphasic Personality Inventories underestimate booklet-based scores? *Journal of Clinical Psychology*, **48**, 744-748.

(Manuscript received September 21, 1998;
revision accepted for publication February 3, 1999.)