Psychological Assessment
2000, Vol. 12, No. 1, 3–5

DOI: 10.1037//1040-3590.12.1.3

Introduction to the Special Section on the Use of Computers for Making Judgments and Decisions

Howard N. Garb
Pittsburgh Veterans Affairs Medical Center and University of Pittsburgh

Computers have the potential to greatly increase the accuracy of judgments and decisions that are made in the area of psychological assessment. The validity of present-day computer programs are described in the first 2 articles of this Special Section. Both computer-based test interpretation programs and mechanical prediction rules (e.g., statistical prediction rules) are described. Limitations of present-day computer programs are described in the 3rd article along with methodological recommendations for building more powerful rules. In the 4th article, a statistical analysis (neural network modeling) is described that may allow us to build better rules.

> As our Modern Age ceases to be modern and becomes an episode of history, our times may well be classed as the later Middle Ages. For while we say time marches forward, all things in time move backward toward the middle and eventually to the beginnings of history.
>
> (Bishop, 1968, p. 11)

To improve psychological assessment (that is, to bring psychological assessment into a new Age), it is necessary to do more than construct new tests and structured interviews: We also need to improve the way that assessment information is used to make judgments and decisions. Recommendations have been made for improving the way clinicians make judgments (e.g., Garb, 1998), but the purpose of this Special Section is to describe some of the latest research and most promising leads on using computers to make judgments and decisions.

Automated-assessment computer programs (also called computer-based test interpretation programs) are already widely used to interpret test results. For example, Raymond Fowler estimated that by 1985 automated-assessment programs had been used to interpret the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942) results for over 1.5 million clients in the United States (Fowler, 1985). Automated-assessment computer programs consist of a series of if–then statements that are written by expert clinicians. For example, if the *F* scale on the MMPI-2 is 11 points or higher than the *K* scale, then a report generated by a computer program may note that the client may be malingering. Though automated-assessment programs have been widely used, their validity has been challenged (Adams & Heaton, 1985; Matarazzo, 1986; Moreland, 1985). Furthermore, guidelines for studying their validity have frequently been ignored (Snyder, Widiger, & Hoover, 1990). In my recent book, *Studying the Clinician: Judgment Research and Psychological Assessment* (Garb, 1998), I presented an updated review and concluded that "most automated-assessment programs have not been shown to be valid" (p. 229). For the first article of this Special Section, James Butcher, the author of one of the most widely used automated-assessment programs, was invited to describe research that supports the use of automated-assessment programs.

The article by Butcher, Perry, and Atlis (2000) is composed of two main parts. After a brief introduction, they compare the computer administration of tests to the paper-and-pencil administration of tests. That is, they review research on whether different results are obtained depending on whether a client takes a test by reading and answering questions on a computer terminal or by reading a test booklet and marking answers with a pencil. Also in this section, they compare computerized interviews to interviews conducted by clinicians. The comparison of computer administered tests and paper-and-pencil administered tests is important because when clinicians interpret test results, they frequently make use of research results that were obtained by giving clients paper-and-pencil administered tests. Butcher et al. conclude that computer administered tests and paper-and-pencil administered tests yield similar results, but most of this research is limited to the use of the MMPI and MMPI-2. The comparison of computerized interviews with interviews conducted by clinicians is important for a different reason: if they yield comparable results, then the use of computerized interviews might become widely adopted. Butcher et al. review only a few studies on this topic: The results indicate that computerized interviews and clinical interviews yield results that are significantly different.

The next section of Butcher et al.'s article is more closely tied to the theme of this Special Section. In this section, they describe the validity of automated assessment programs, including programs for personality assessment (e.g., for interpreting MMPI results), psychodiagnosis, and neuropsychological assessment. Surprisingly, with regard to the validity of computer-based test reports in the area of personality assessment, Butcher et al. were able to find only four studies that had been conducted during the 1990s. Negative findings were reported in two of the studies (Guastello & Rieke, 1990; Prince & Guastello, 1990), and in a third study (Eyde, Kowal, & Fishburne, 1991) "the overall results

Howard N. Garb, Department of Behavioral Health, Pittsburgh Veterans Affairs Medical Center, and Department of Psychiatry, University of Pittsburgh.

Correspondence concerning this article should be addressed to Howard N. Garb, Department of Behavioral Health (116A-H), Pittsburgh Veterans Affairs Medical Center, 7180 Highland Drive, Pittsburgh, Pennsylvania 15206-1297. Electronic mail may be sent to garb.howard@pittsburgh.va.gov.

indicated that the validity of the narrative outputs varied" (Butcher et al., 2000, p. 10). In the fourth study (Butcher et al., 1998), Butcher et al. (2000; this issue) reported that "In 87% of the reports, at least 60% of the computer-generated narrative statements were believed to be appropriate" (p. 18). Butcher et al. (2000; this issue) believe that this result supports automated assessment, but implicit in their statement is the finding that as many as 40% of the computer-generated narrative statements were less than appropriate in 87% of the reports. In addition to describing results for personality assessment, Butcher et al. describe results for psychodiagnosis and neuropsychological assessment: No evidence is given that would indicate that the programs are as accurate as, or more accurate than, clinicians.

In contrast with automated assessment programs, statistical prediction rules are mathematical equations (e.g., linear regression equations) that are usually empirically based (parameters and weights are usually derived from empirical data; unit weight linear models are an exception). Because statistical prediction rules are usually empirically based, they have done well when compared with clinicians and automated assessment programs (Garb, 1998). Strong advocates of statistical prediction rules have been critical of "the failure of mental health professionals to apply in practice the strong and clearly supported empirical generalizations demonstrating the superiority of actuarial over clinical prediction" (Meehl, 1986, p. 370, abstract; also see Dawes, Faust, & Meehl, 1989). In the second article of this Special Section, Grove, Zald, Lebow, Snitz, and Nelson (2000) present the results of a meta-analysis on clinical versus mechanical prediction. Their use of the term *mechanical prediction* refers to the use of explicit mathematical formulas, actuarial tables, and other algorithmic procedures. Their article is of interest because they include twice as many studies as have been included in previous reviews and because this is the first meta-analysis conducted on this topic.

For their meta-analysis, Grove et al. (2000) included studies from the areas of medicine, psychiatry, clinical and counseling psychology, personnel psychology, and other fields related to the assessment and prediction of human behavior. In all of the studies, clinical judgments were compared with mechanical predictions. In most of the 136 studies, mechanical prediction rules were as accurate as, or more accurate than, human judges. In fact, on average, statistical prediction rules were about 10% more accurate than clinical judges. Overall, these results indicate that computers should be used to make judgments and decisions.

The utility of statistical prediction rules varies as a function of judgment task. In the third article of this Special Section, I describe the validity of present-day statistical prediction rules and make recommendations for building new rules. Results have been disappointing for the following tasks: diagnosis of mental disorders, describing personality traits and psychiatric symptoms, predicting suicide, and the diagnosis and description of neurological impairment. Furthermore, statistical prediction rules have rarely been used to make causal judgments or treatment decisions (though important exceptions exist in the area of behavioral assessment). Perhaps the best statistical prediction rules involve the prediction of violence (e.g., Gardner, Lidz, Mulvey, & Shaw, 1996) and the drawing of inferences about functional relationships among antecedents, behaviors, and consequences in behavioral assessment (Schlundt & Bell, 1987; also see Shiffman, 1993). Though most present-day statistical prediction rules are of limited value, innovative strategies can be used to build a new generation of statistical prediction rules. Recommendations are made for using innovative strategies to obtain criterion scores and identify optimal predictors.

Advances in statistical prediction will occur not only with the use of new methods for obtaining criterion scores and identifying optimal predictors, but also with advances in statistical analyses. Promising analyses include neural network modeling and taxometric analyses. Neural network modeling is described in this Special Section (Price et al., 2000, this issue), but taxometric analyses are not, in part because they have been described at length elsewhere (e.g., Meehl, 1995; Waller & Meehl, 1998). Using taxometric analyses, one can try to determine if it is appropriate to say that individuals lie along a continuum or can be sorted into categories. According to Meehl (1995, p. 266, Abstract), "Revision of diagnostic systems should be based on taxometric analysis rather than on committee decisions based on clinical impression and nontaxometric research." Thus, in the future, diagnoses may be made by using rules that are derived by conducting taxometric analyses. However, problems with using taxometric analyses have been described (Miller, 1996; though also see Meehl, 1996) and attempts to classify patients using taxometric analyses have received only mixed support in the area of psychodiagnosis, though improvements in the analysis continue to be made (Garb, 1996).

The article by Price et al. (2000) introduces neural network models to the area of psychological assessment. Neural network models (also called artificial neural networks) are computing systems that mimic how the brain works (Rumelhart, McClelland, & the PDP Research Group, 1986). They have become increasingly important for many fields including cognitive neuroscience (to model how the brain works) and computer science (for solving recognition tasks, image and signal processing tasks, and other tasks that require artificial intelligence). Neural network models are powerful because of their ability to detect complex (i.e., higher order) relationships in a data set. In their article, Price et al. compare artificial neural networks to linear and quadratic discriminant analyses and to logistic regression analyses. They use this comparison for two data sets, and for both data sets the artificial neural networks are more accurate than are the other statistical rules.

It should be possible to use neural network models for a range of tasks in psychological assessment. For example, Price et al. (2000) used artificial neural networks to predict mortality for a sample of Vietnam veterans, drug-using veterans, and civilians (the data are from the Washington University Vietnam Era Study; Robins, 1974). Psychologists are not normally interested in predicting mortality, but certainly the same analyses could be used to predict other behavioral outcomes (e.g., suicide, violence, relapse). Price et al. also used artificial neural networks to make diagnoses of mental disorders. The data they analyzed had been collected as part of the Epidemiologic Catchment Area project, one of the largest epidemiologic psychiatric studies ever conducted in the United States (Robins & Regier, 1991). When given the symptom ratings that were collected for this project, artificial neural networks were more accurate than traditional statistical analyses in predicting diagnoses that had been made by using the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–III;* American Psychiatric Association, 1980). This finding has little clinical significance (why should clinicians use artificial neural networks to predict diagnoses that can more easily be made

by using *DSM?*), but it demonstrates that artificial neural networks can be used to uncover complex interactions. In future studies, artificial neural networks can be used to predict diagnoses that are made after collecting data which are not usually available in clinical practice (e.g., diagnoses that are made after collecting longitudinal data).

Finally, the concluding article of this Special Section was written by Snyder (2000). Snyder was invited to comment on the other articles.

In conclusion, computers will transform psychological assessment. Present-day automated assessment programs and statistical prediction rules are of limited value, but changes in the way that statistical prediction rules are derived and evaluated and the use of innovative statistical analyses will lead to the appearance of powerful rules for making judgments and decisions.

References

Adams, K. M., & Heaton, R. K. (1985). Automated interpretation of neuropsychological test data. *Journal of Consulting and Clinical Psychology, 53,* 790–802.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Bishop, M. (1968). *The horizon book of the Middle Ages.* New York: American Heritage.

Butcher, J. N., Berah, E., Ellersten, B., Miach, P., Lim, J., Nezami, E., Pancheri, P., Derksen, J., & Almagor, M. (1998). Objective personality assessment: Computer-based MMPI-2 interpretation in international clinical settings. In C. Belar (Ed.), *Comprehensive clinical psychology: Sociocultural and individual differences.* New York: Elsevier.

Butcher, J. N., Perry, J. N., & Atlis, M. M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment, 12,* 6–18.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989, March 31). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Eyde, L., Kowal, D. M., & Fishburne, F. J. (1991). The validity of computer-based test interpretations of the MMPI. In T. B. Gutkin & S. L. Wise (Eds.), *The computer and the decision-making process* (pp. 75–123). Hillsdale, NJ: Erlbaum.

Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology, 53,* 748–759.

Garb, H. N. (1996). Taxometrics and the revision of diagnostic criteria. *American Psychologist, 51,* 553–554.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996). Clinical versus actuarial predictions of violence by patients with mental illnesses. *Journal of Consulting and Clinical Psychology, 64,* 602–609.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19–30.

Guastello, S. J., & Rieke, M. L. (1990). The Barnum effect and validity of computer-based test interpretations: The Human Resource Development Report. *Psychological Assessment, 2,* 186–190.

Hathaway, S. R., & McKinley, J. C. (1942). *The Minnesota Multiphasic Personality Inventory.* Minneapolis: University of Minnesota Press.

Matarazzo, J. D. (1986). Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma. *American Psychologist, 41,* 14–24.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275.

Meehl, P. E. (1996). MAXCOV pseudotaxonicity. *American Psychologist, 51,* 1184–1186.

Miller, M. B. (1996). Limitations of Meehl's MAXCOV-HITMAX procedure. *American Psychologist, 51,* 554–556.

Moreland, K. L. (1985). Validation of computer-based test interpretations: Problems and prospects. *Journal of Consulting and Clinical Psychology, 53,* 816–825.

Price, R. K., Spitznagel, E. L., Downey, T. J., Meyer, D. J., Risk, N. K., & El-Ghazzawy, O. G. (2000). Applying artificial neural network models to clinical decision making. *Psychological Assessment, 12,* 40–51.

Prince, R. J., & Guastello, S. J. (1990). The Barnum effect in a computerized Rorschach interpretation system. *Journal of Psychology, 124,* 217–222.

Robins, L. N. (1974). *The Vietnam drug user returns* (Special Action Office Monograph, Series A, No. 2). Washington, DC: U.S. Government Printing Office.

Robins, L. N., & Regier, D. A. (Eds.). (1991). *Psychiatric disorders in America.* New York: Free Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations into the microstructure of cognition, Volumes 1 and 2.* Cambridge, MA: MIT Press.

Schlundt, D. G., & Bell, C. (1987). Behavioral assessment of eating patterns and blood glucose in diabetes using the Self-Monitoring Analysis System. *Behavior Research Methods, Instruments, & Computers, 19,* 215–223.

Shiffman, S. (1993). Assessing smoking patterns and motives. *Journal of Consulting and Clinical Psychology, 61,* 732–742.

Snyder, D. K. (2000). Computer-assisted judgment: Defining strengths and liabilities. *Psychological Assessment, 12,* 52–60.

Snyder, D. K., Widiger, T. A., & Hoover, D. W. (1990). Methodological considerations in validating computer-based test interpretations: Controlling for response bias. *Psychological Assessment, 2,* 470–477.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures.* Thousand Oaks, CA: Sage.

Received December 2, 1998
Revision received February 26, 1999
Accepted March 2, 1999 ■