— PROGRAM ABSTRACTS/ALGORITHMS —

CASPER: A personal computer package for exploring psychometrics

WILLIAM V. CHAMBERS

University of South Florida, Fort Myers, Florida
and

JAMES W. GRICE

University of New Mexico, Albuquerque, New Mexico

CASPER is a psychometrics software package suitable for instructional and research applications with IBM-PC-compatible computers. CASPER lets the user simulate or directly enter psychometric data. Numerous statistical analyses, file handling procedures, and graphics are included. Analyses include factor analysis, multiple regression, correlation/partial correlation, moments analysis, reliability analysis, and item analysis.

Psychometric methods are an important part of every psychologist's undergraduate and graduate education. Tests and measurements courses are taught in most psychology and education departments. Although many computer programs have been developed to teach the statistics used in experimental design, few or none have been developed that specifically address psychometric statistics. The Construct Analysis and Simulation package for Education through Research (CASPER), an integrated, easy-to-use package that requires no programming skills, was written to fill this void.

CASPER was designed to allow students and researchers to explore psychometrics through both Monte Carlo simulations and real data analysis. Users simulate true scores, observation scores, and measurement error, using random numbers or limiting case data, in order to model psychometric construct domains. These domains may be constituted from either interval or nominal scales, allowing the simulation of Likert scale, true-false, or multiple-choice tests.

The logic of the simulation follows the rationale of the classical true score model (Nunnally, 1978). The user may vary the number, range, and distributions of true scores. Errorless item scores are then generated; these are the ratings applied to questionnaire items if no measurement error is involved. Total scores derived from errorless item scores are identical to true scores, and the ratings are perfectly valid and reliable. To make the data more realistic, users may add various types and amounts of error variance to the errorless item scores, creating observation

scores. The observation scores constitute a sample from the hypothetical construct domain. By systematically varying the parameters used to define domains, and by observing the consequences disclosed through psychometric analysis of the observation scores, the user can study the logic of the classical true score model. The user thus discovers psychometric principles through actively manipulating the data rather than through passive memorization.

Input. CASPER is menu-driven. Data are simulated, entered through the keyboard, or loaded from ASCII files. Data sets may contain up to 30 items (k) and 300 subjects (n). Extensive editing and filing procedures allow considerable flexibility for the modification of data. Data for particular items or subjects may be reflected, deleted, edited, standardized, or dichotomized. Data sets and results may be written to screen or saved to disk. Different domains can be concatenated to simulate the multi-dimensional structures reflected in multitrait, multimethod matrices. A domain specification option reminds the user of the current domain's parameters and provides summaries for organizing the laboratory notebooks. A set of defaults in the construct domain simulation option allows one to generate data by simply pressing the return key. These defaults walk the new user through the simulation process, demonstrating the method and rationale as the user progresses. Options to view or print the data are provided at each stage of the simulation.

Output. CASPER provides an assortment of analytical procedures. The factor analysis procedures include principal components analysis based on the Jacobi method (Harmon, 1967); principal axis analysis, with or without iterative commonality estimation (Gorsuch, 1983); Guttman's image analysis (Nunnally, 1978); multiple group confirmatory analysis (Gorsuch, 1983); and raw score factor analysis (Nunnally, 1978). Varimax rotation is performed, and component or factor scores are available (Harmon, 1967). The results may be saved for further analysis or plotted as factor loadings or scree plots.

CASPER incorporates a moments analysis (Ferguson, 1959) that provides measures of the mean, standard deviation, skewness, and kurtosis. Measures of association included are correlations, partial correlations, variances, and covariances, as well as Euclidean distances. An item analysis of multiple-choice data is available, which includes frequency, error, difficulty, and discrimination indices (Crocker & Algina, 1987). A multiple regression analysis that provides predicted scores and sums of squares analysis is also included (Pedhazur, 1982).

Reliability analysis procedures include Cronbach's alpha (Cronbach, 1984), Hoyt's ANOVA (Crocker & Al-

Requests for reprints should be sent to William V. Chambers, Department of Psychology, University of South Florida, 8111 College Parkway, Fort Myers, FL 33919.

gina, 1987), *rkk* (Nunnally, 1978), and split-half measures incorporating Spearman-Brown corrections (Gulliksen, 1950) and Rulon's method (Crocker & Algina, 1986). Reliabilities of longer or shorter tests can be estimated from a Spearman-Brown procedure (Gulliksen, 1950), as can the length of tests needed to reach specified levels of reliability. A correction for attenuation option is also included (Ghiselli, Campbell, & Zedeck, 1981).

Accuracy. All of the analyses have been checked against results given in the cited texts and/or with SAS or SPSS analyses.

Graphics. In addition to histograms, scree plots, scatter plots, and factor loading plots, CASPER includes a graphics procedure designed to illustrate the consequences of error and number of items (k) on validity and reliability. The rationale of this procedure follows the logic of point estimation discussed by Marascuio (1971). The procedure essentially teaches the concepts of accuracy and precision.

Validity and reliability are illustrated as problems in point estimation by plotting item/observation scores and total scores against a targeted domain. A large circle representing the target domain is drawn on the screen. Small circles are progressively displayed, each representing the addition of a new questionnaire item. Medium-sized circles are used to reflect the current status of the total scores. Total score coordinates are calculated by averaging all current item scores each time k is incremented by one item. The user determines the level of validity (average accuracy) and reliability (precision). By systematically varying these parameters, users observe total scores stabilize with higher numbers of items (k), the spread of item scores broadens as reliability error is increased, and the proximities of total scores vary from the target center as validity error is increased.

Exploration strategies. CASPER can be used to analyze either empirical or simulated data. This makes possible an array of instructional strategies. Easy simulation of psychometric data, however, is probably the major advantage of CASPER over other statistical packages.

Generally, the first step in data simulation is the creation of true scores. The user may simulate true scores from random numbers, varying their range and distribution. Alternatively, hypothetical true scores that represent limiting cases may be entered through the keyboard or from files. Examples of such limiting cases include data sets in which subjects are perfectly differentiated into two or more groups by their true scores. An interesting contrast to such data is the case in which all subjects have the same true scores. In this case, there would be no differentiation between subjects. Similar modeling of true scores can produce a continuum of ability, with groups of subjects being distributed across numerous points on a scale. By altering the distributions of such groups across the range of true scores, the user can simulate many different latent trait models.

After true scores have been created, item scores can be generated either by the computer or by direct entry. CASPER allows the user to select the type of response scale being simulated and to place various constraints on the distribution of simulated responses. The user may also enter limiting case data that transform the true scores into a questionnaire format. If the user is analyzing data collected empirically—for example, from actual questionnaires collected in the laboratory—the data can be directly entered as item scores. In this case, the user would probably chose to skip over the true score option altogether, since the true scores are not known.

When the user is not simulating limiting cases or analyzing actual empirical data, item scores are usually further modified by the addition of simulated error variance. These scores are referred to as *observation scores*. In order to simulate the error that converts errorless items scores to observation scores, the user first defines a range and distribution of random numbers. These numbers then are added either to all the item scores or to particular item scores. The extent and distribution of measurement error can thus be assigned to make particular items differ in reliability.

After observation scores have been generated, the data may be saved to disk, and a variety of statistical analyses can be performed. The type of analysis generally depends on the research question. Some questions addressed by students as they have completed projects with CASPER have included: (1) How does increasing the reliability error influence Cronbach's alpha? (2) When will alpha be lower than the odd/even split-half reliability, and how can factor analysis be used to clarify this condition? (3) When are factor loadings made more representative by varimax rotation? (4) Do factor scores improve multiple regression predictions of true scores? (5) Are normal distributions of total scores more likely than uniform distributions, when true scores are uniform? (6) Can the alteration of correlation matrix values create negative eigenvalues? (7) How accurate are the Spearman-Brown estimates of reliability for increased k ? (8) Does the index of discrimination detect poor items as well as the correlation of the items with total scores does? (9) What is the relationship between total scores and factor scores? (10) What happens when principal component loadings are entered as multiple group weights? There are many more examples, but these illustrate the depth of the questions that students have asked and answered with CASPER.

System requirements. CASPER operates on PC-compatibles with 640K, one disk drive, and a graphics card. A hard disk or second disk drive is helpful for saving and reading data files.

Availability. A free copy of the compiled version of CASPER is available from William V. Chambers. Users are encouraged to share copies with others, but commercial use or sale of the program by the user is expressly forbidden. To obtain a copy of CASPER, make a request

for the package on your department stationary, send a *formatted* double-sided, double density 5.25-in. or 3.5-in. diskette, and include a self-addressed, stamped package for its return.

REFERENCES

- CROCKER, L., & ALGINA, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- CRONBACH, L. J. (1984). *Essentials of psychological testing*. New York: Harper & Row.
- FERGUSON, G. A. (1959). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- GHISELLI, E. E., CAMPBELL, J. P., & ZEDECK, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: W. H. Freeman.
- GORSUCH, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
- GULLIKSEN, H. (1950). *Theory of mental tests*. New York: Wiley.
- HARMON, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.
- MARASCULIO, L. A. (1971). *Statistical methods for behavioral science research*. New York: McGraw-Hill.
- NUNNALLY, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- PEDHAZUR, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.

(Manuscript received August 3, 1990;
revision accepted for publication December 21, 1990.)

Notices and Announcements

PSYCOLOQUY

An experimental "BBS of the air" has been created under the sponsorship of the Science Directorate of the American Psychological Association. PSYCOLOQUY is an international, interdisciplinary electronic network for BBS-style "scholarly skywriting."

Contributors are invited to post brief squibs (preferably not much more than a screenful, although very short articles are potentially acceptable, too) reporting recent ideas or findings on which the author wishes to invite "skywriting" discussion from the world psychological community. PSYCOLOQUY provides an excellent means of "test-piloting" material that may eventually become a target article for open peer commentary in BBS.

All contributions are refereed by members of PSYCOLOQUY's Editorial Board, which covers all areas of psychology and related fields.

Subscriptions to PSYCOLOQUY are free, but the "subscriber" must have an electronic mail ("e-mail") address. To subscribe, send the following one-line message to listserv@tcsvm.bitnet

sub psych Lastname Firstname

(substituting the subscriber's last name and first name).

After signing on, submissions can be posted to psych@tcsvm.bitnet