20 Software Packages for Assessing Test Dimensionality
(Version: November 15, 2007)

Nina Deng and Ronald Hambleton
University of Massachusetts Amherst

In this brief report, 20 software packages for assessing test dimensionality are briefly described. When possible, we have provided a citation for more information and/or a website for follow-up investigations.

A FORTRAN program assessing unidimensionality of binary data
Author: R. Nandakumar
Source: Multivariate Behavioral Research, 1993, 28(1), 63-66.
Features:

Outlines Holland and Rosenbaum's (1986) methodology to assess unidimensionality.

ALSCAL
Author: Forrest W. Young
Sources: L. L. Thurstone Psychometric Laboratory http://forrest.psych.unc.edu/research/alscal.html
SPSS Inc. http://www.spss.com/
SAS Institute, Inc. http://www.sas.com/
Features:

Perform metric or nonmetric Multidimensional Scaling (MDS) and Unfolding with individual differences options. Analyze one or more matrices of dissimilarity or similarity data.

Nonmetric multidimensional scaling solutions had relatively lower stress. (+) Solution based on matrix of original dissimilarity were more closely related to the underlying configurations. (-) determination of dimensionality by inspection of the stress plot was somewhat more difficult for solutions based on (Dragsgow, F., & Jones, L. E., 1979).

(+) The PC statistic correctly identified the dimensionality of the unidimensional data, (-) whereas the use of Euclidean distances suggested the two-parameter unidimensional data were multidimensional. (+) Both procedures correctly identified multidimensionality under the low correlation conditions. (-) Unable to detect multidimensionality when the dimensions were highly correlated (Meara, K., Robin, F., & Sireci, S. G., 2000).

(+) MDS and confirmatory factor analysis (CFA) were able to correctly identify the number of latent dimensions (Ayala, R. J. D., & Hertzog, M. A., 1991).

(+) A viable methodology in assessing the dimensionality of ordered test or attitude data (De Ayala, R. J., & Hertzog, M. A., 1989, March).

AMOS: Analysis of Moment Structures
Author: J. L. Arbuckle
Source: SPSS Inc. http://www.spss.com/amos16/
Features:

(+) Estimate models with missing data and polytomous data.

CASPER
Authors: William V. Chambers, & James W. Grice
Source: Educational and Psychological Measurement, 1992, 52(3), 599-600.
Features:

Model psychometrics and factor analysis and was designed for teaching in tests and measurements courses. It can simulate data and tests, do different factor analyses including principle components analysis, principle axis analysis, image analysis, and multi-group confirmatory analysis. Multiple regression, partial correlation, reliability analyses, an item analysis, and various plotting procedures are also provided.

CHIDIM
Authors: A. F. De Champlain & K. L. Tang
Source: Educational and Psychological Measurement, 1997, 57(1), 174-178.
Features:

A fortran program assessing the dimensionality of binary item responses based on McDonald's nonlinear factor analytic model. In particular, an approximate chi-square statistic was introduced to be accurate in identifying the true number of underlying dimensions.

DETECT
Authors: J. Douglas, H. R. Kim, L. Roussos, W. F. Stout, & J. Zhang
Source: Assessment System Corporation
http://assess.com/xcart/product.php?productid=385
Features:

Acronym for Dimensionality Evaluation To Enumerate Contributing Traits, a nonparametric dimensionality assessment procedure using estimated item pair conditional covariances. Operate in either a confirmatory or an exploratory mode. Presumes dichotomous item scoring. Currently incorporated into DIMPACK with Windows interface.

(+) Most useful when the data display approximate simple structure. (-) Not effective for complex structure or high correlated dimensions.

(+) Accurately and consistently cluster items when as many as 30% of the items display complex structure, the correlation between dimensions is less than or equal to 0.75 and the sample size is at least 1,000. (+) Or 50% of the items

displaying complex structure, and correlation less than or equal to 0.60 and the sample size is at least 1,000. (-) Not work well when the correlation is or larger than 0.90. (Gierl, Leighton, & Tan, 2006).

(-) Exhibit statistical bias in the large-sample estimation of the item-pair conditional covariances when multidimensionality is present. (+) The bias was negligible with 20 or more items (Roussos, & Ozbek, 2006).

DIMTEST

Authors: William Stout and Ratna Nandakumar

Source: Assessment System Corporation
http://assess.com/xcart/product.php?productid=385

Features:

A nonparametric testing procedure that assesses lack of latent unidimensionality for a dichotomously scored test. Assess the statistical significance of the possible dimensional distinctiveness between two specified subtests: the Assessment Subtest (AT) and the Partitioning Subtest (PT). Operate in either a confirmatory or an exploratory mode. Currently incorporated into DIMPACK.

(-) sensitive to whether the multidimensional data arose from a compensatory model or a partially compensatory model (Hattie, Krakowski, Rogers, & Swaminathan, 1996).

(+) Very low type I error with no presence of guessing. Comparatively lower type I error with guessing than NOHARM (Finch & Habing, 2007).
(-) Lowe power and high type I error with small sample size; underestimate the dimensionality with multidimensional data.

(-) some trouble detecting multidimensionality when the number of test items was small. (Meara, Robin, & Sireci, 2000).

(+) show excellent power in detecting a lack of unidimensionality over HR' method and non-linear factor analysis. (Nandakumar, 1991).

(+) identify for noncompensatory multidimensionality when the sample size was large and the interability correlation was low. (-) fail to reject the hypothesis of essential unidmensionality for compensatory multidimensional data. With greater relative influence of the second dimension, the rejection power increased when test length and sample size increased, and when interability correlation decreased (Deng, Ansley, 2000??)

(-) the power of DIMTEST decreased as the mean of the secondary ability distribution approached the extremes and/or as the standard deviation of the secondary ability distribution approached zero. (Walker, Azen, & Schmitt, 2006).

HCA/CCPROX

Authors: W. Stout, B. Habng, J. Douglas, H. R. Kim, L. Roussos, & J. Zhang
Source: Assessment System Corporation
http://assess.com/xcart/product.php?productid=230
Features:

Stands for agglomerative hierarchical cluster analysis (HCA) using the unweighted pair-group method of averages with Roussos' proximity measure CCPROX based on the item pair covariances conditioned on the remaining items.

(+) perform on either dichotomous or polytomous scored items.
(+) correctly partition the test into dimensionally homogeneous item clusters even for very high correlations between the latent dimensions, when approximate simple structure holds (Roussos, Stout, & Marden, 1998).

EQS
Author: P.M. Bentler
Source: Multivariate Software, Inc.
http://www.mvsoft.com/products.htm
Features:

Conduct structural equations models including multiple regression, multivariate regression, confirmatory factor analysis, structured means analysis, path analysis, and multiple population comparisons. Provide the analysis on data that may not be multivariate normally distributed with the Satorra-Bentler scaled chi-square, robust standard errors, and the Yuan-Bentler distribution-free statistics.

(+) Estimate models with missing data.

(+) Iterative procedure improves confirmatory model fit by removing the poorest fitting items as revealed by a variable fit index (Hofmann, 1995).

LDIP
Authors: Seock-Ho Kim, Allan S. Cohen, and Yuan-Horng Lin
Source: Applied Psychological Measurement, 2006, 30(6), 509-510.
Features:

Provide indices of local dependence for polytomous items from the graded response and generalized partial credit models from either MULTILOG or PARSCALE.

The indices include the Pearson chi-square statistic $\chi 2$, the likelihood ratio chi-square statistic G2, Yen's index of local dependence Q3, and the Fisher-transformed correlation difference statistic Zd.

LISREL
Authors: K. G. Jöreskog and D. Sörbom
Source: Scientific Software International, Inc.
http://www.ssicentral.com/lisrel/new.html
Features:

A program for structural equation modeling for both dichotomous and polytomous data. Also incorporate statistical applications including PRELIS for data manipulations and basic statistical analyses; MULTILEV for hierarchical linear and non-linear modeling; SURVEYGLIM for generalized linear modeling, CATFIRM for formative inference-based recursive modeling for categorical response variables; CONFIRM for formative inference-based recursive modeling for continuous response variables; MAPGLIM for generalized linear modeling for multilevel data.

(+) Estimate models with polytomous data and randomly missing data.

(+) More than chi-square index is provided
(+) High rejection rates for multidimensional data. (-) Inflated type I error with chi-square fit statistic for unidimensional data. (-) Influenced by conditions manipulated with small sample sizes and short test lengths (De Champlain, 1998).

(-) Higher type I error with chi-square goodness-of- fit statistic for unidimensional data in LISREL 8/PRELIS 2 than in LISREL 7/PRELIS 1 (De Champlain, 1997).

(-) Not work well with sample size less than 500 (De Champlain, 1998).

MICROFACT
Author: N. G. Waller
Source**:** Assessment Systems Corporation
http://assess.com/xcart/product.php?productid=241&cat=0&page=1
Features:
A factor analysis program with Windows interface for mainframe size problems for both dichotomous and ordered polytomous data. A demo version is free to download with a maximum of 30 variables.

MOKSCAL
Authors: Johannes Kingma & Terry Taerum
Source: Applied Psychological Measurement, 1988, 12(2), 188.
Features:
Analyze the Mokken (1971) scale analysis based on a nonparametric item response model (Mokken & Lewis, 1982)

Test the assumption of double monotony. Compute three coefficients of scalability and check monotone of homogeneity. Computes four different coefficients of reliability and the biserial correlations.

Mokken Scaling Program (MSP)
Authors: I .W. Molenaar, P. Debets, K. Sijtsma, & B. T. Hemker.
Source: Assessment Systems Corporation
http://assess.com/xcart/product.php?productid=242&cat=30&page=1

Features:

Evaluate model fit of a given scale or construct one or more unidimensional scales from an item pool using nonparametric cumulative item response theory. Check the assumptions of monotone homogeneity and double monotonicity. Assess test reliability and detect misfitting items. Analyze either dichotomous or polytomous data.

(-) require much lower correlations before identifying distinct dimensions than DETECT and parameter cluster analysis (Mroch, & Bolt, 2006).

(-) sequential item selection procedure may not find the dominant underlying dimensionality. Hierarchical clustering methods can improve the searching for the dominant dimensionality (van Abswoude, Vermunt, & Hemker, van der Ark, Andries, 2004??).

(+) The average within-scale Hi combined with a stochastic nonhierarchical clustering algorithm was successful in reflecting the underlying dimensionality (van Abswoude, Vermunt, & Hemker, 2007).

Mplus*: Statistical analysis with latent variables*
Authors: B. O. Muthen & L. K. Muthen
Source: http://www.statmodel.com/index.shtml
Features:

A statistical modeling program analyzes cross-sectional and longitudinal data, single-level and multilevel data with continuous, censored, binary, ordered categorical (ordinal), and unordered categorical (nominal) observed variables. Have features for missing data, complex survey data, and Monte Carlo simulation studies. Can analyze a matrix of tetrachoric correlations for dichotomous variables.

(+) Estimate models with polytomous data and missing data.

(+) When data assume guessing, the proportion of variance and the RMSR reduction indices more accurately estimated dimensionality in Mplus than in TESTFACT (Yeh, 2007)

NOHARM
Authors: Colin Fraser, & R. P. McDonald
Source: Downloadable from http://kiptron.usc.edu/software.html
Features:

An IBM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory based on theory developed by R. P. McDonald.

(-) Only handling dichotomous data.

(+) Very low type I error with no presence of guessing. (-) Unacceptably high type I error when there is guessing (Finch, & Habing, 2007).

(+) The performance of the approximate C2 based on McDonald's (1967) nonlinear factor analytic was as good as Stout's T statistic in all conditions and was better than Stout's T statistic with smaller sample sizes and shorter tests (Gessaroli, & De Champlain, 1996).

Perform comparably to DETECT in simultaneously finding the correct number of dimensions and clustering items correctly. (+) generally better able to determine the number of underlying dimensions, (-) but less able to group items together than DETECT. (-) More likely to incorrectly groups them together, while DETECT is more likely to incorrectly separate items. (Finch, & Habing, 2005).

Poly-DIMTEST
Authors: H. Li and W. Stout
Source: Assessment Systems Corporation
http://assess.com/xcart/product.php?productid=249&cat=0&page=1
Features:

A modified version of DIMTEST that can be used to assess lack of latent unidimensionality for polytomously scored item response data

(+) correctly confirm unidimensionality for unidimensional simulated data, with observed level of significance slightly below the nominal level. (+) effectively detect lack of unidimensionality in various two-dimensional tests.
Power increased as the sample size and test length increased, and the correlation between the Os decreased.
Pearson correlations to select ATI items led to equally good or better performance than using polychoric correlations.
(Nandakumar, Yu, Li, & Stout, 1998)

(+) T-empirical Type I error were at or near nominal values for the generated data (+) and unaffected by the manipulation of sample size. (-) lack the power with fewer than 20 items (De Champlain, Gessaroli, Tang, & De Champlain, J. 1998).

TESTFACT
Authors: D. Wilson, R. Wood, S. Schilling, & R. Gibbons
Source: Scientific Software International, Inc.
http://www.ssicentral.com/irt/index.html
Features:

Implements full-information factor analysis of inter-item tetrachoric correlations, and methods of factor analysis based on item response theory (IRT).

(+) Estimate models with unreached, omitted, or not-presented data, data with guessing, and polytomous data. (+) Conduct bifactor analysis.

(+) Correct for the guessing effect and omitted items. Have option of imposing constraints on item parameter estimates in maximum likelihood factor analysis to avoid the Heywood case (Muraki, 1984).

(+) The chi-square test and parallel analysis performed better with TESTFACT than with Mplus when data assume guessing. (+)more consistency in the dimensionality assessment with TESTFACT than with Mplus (Yeh, 2007).

(-) Inflated type I error for unidimensional data with G2 difference test, when the sample sizes are small and tests are short. (-) influenced by the conditions manipulated (De Champlain, 1998).

TESTSIM

Authors: W.F. Stout, R. Nandakumar, B. Junker, H. H. Chang, & D. Steidinger
Source: University of Illinois, Department of Statistics, Urbana IL.

TETRAD

Authors: Clark Glymour, Richard Scheines, Peter Spirtes, Kevin Kelly

Sources: Academic Press, Inc

Multivariate Behavioral Research, 1988, 23(2), 279-280.

Features:

Apply heuristic search techniques to aid in the specification of structural equation models.

(+) successfully confirmed one dimension in the single-construct data set and confirmed two dimensions in the combined data set, (-) yet excluded one item from each cluster, for no obvious reasons. (Yu, Ho, Popp, Osborn, DiGangi, Samuel, & Jannasch-Pennell, 2007??).