DOCUMENT RESUME

ED 431 792	TM 029 878
AUTHOR	Egan, Karla L.; Sireci, Stephen G.; Swaminathan, Hariharan; Sweeney, Kevin P.
TITLE	Effect of Item Bundling on the Assessment of Test Dimensionality. Laboratory of Psychometric and Evaluative Research Report No. 328.
INSTITUTION	Massachusetts Univ., Amherst. School of Education.
PUB DATE	1998-04-00
NOTE	35p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Diego, CA, April 12-16, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Certified Public Accountants; *Evaluation Methods; Licensing Examinations (Professions); Simulation; *Test Items
IDENTIFIERS	*Dimensionality (Tests); *Item Bundles

ABSTRACT

The primary purpose of this study was to assess the effect of item bundling on multidimensional data. A second purpose was to compare three methods for assessing dimensionality. Eight multidimensional data sets consisting of 100 items and 1,000 examinees were simulated varying in terms of dimensionality, inter-dimensional correlation, and number of items loading on each dimension. Analyses were also performed on two samples of examinees from the November 1996 administration of the Uniform Certified Public Accountant examination. The items from both data sets were grouped into bundles that varied in size and content. Principal components factor analysis, maximum likelihood factor analysis, and multidimensional scaling were used to analyze the item bundles as well as the items themselves. Results suggest that item bundling tends to obscure multidimensionality, but analyses on the items themselves overestimate dimensionality. Multidimensional scaling also appeared better able to recover the underlying dimensionality of the data than the other two techniques. (Contains 13 tables and 17 references.) (Author/SLD)

ED 431 792

Effect of Item Bundling on the Assessment of Test Dimensionality[1,2]

Karla L. Egan
Stephen G. Sireci
Hariharan Swaminathan
University of Massachusetts, Amherst

Kevin P. Sweeney
American Institute of Certified Public Accountants

1 Paper presented at the National Council on Measurement in Education, San Diego, CA, April 1998.
2 Laboratory of Psychometric and Evaluative Research Report No. 328. University of Massachusetts at Amherst, School of Education.

Abstract

The primary purpose of this study was to assess the effect of item bundling on multidimensional data. A secondary purpose was to compare three methods for assessing dimensionality. Eight multidimensional data sets consisting of 100 items and 1000 examinees were simulated varying in terms of dimensionality, inter-dimensional correlation, and number of items loading on each dimension. Analyses were also performed on two samples of examinees from the November 1996 administration of the Uniform CPA examination. The items from both data sets were grouped into bundles that varied in size and content. Principal components factor analysis, maximum-likelihood factor analysis, and multidimensional scaling were used to analyze the item bundles as well as the items themselves. Our results suggested that item bundling tends to obscure multidimensionality but analyses on the items themselves overestimate dimensionality. And, MDS appeared better able to recover the underlying dimensionality of the data than the other two techniques.

Introduction

The assessment of test dimensionality remains an important endeavor for psychometricians. Appraising the dimensionality of test response data is necessary for understanding the construct measured by the test and for determining the number of scores needed to adequately summarize test performance. Moreover, tests are increasingly being developed or scored using unidimensional item response theory (IRT) models, which assume the test is measuring a single latent trait (dimension). This increasing use of IRT underscores the need for evaluating test dimensionality to ensure the unidimensionality assumption holds for the data to which these models are applied.

Historically, principal components analysis (PCA) and common factor analysis have been used to evaluate test dimensionality. However, most educational tests contain items that are scored dichotomously, and factor analysis of dichotomous data has been shown to overestimate true dimensionality when the item difficulties are not uniform (Bock, Gibbons, & Muraki, 1988; Green, 1983; Hattie, 1985; McDonald & Ahlawat, 1974). To address this problem, newer techniques have emerged designed specifically for evaluating the dimensionality of dichotomous test data. Among these techniques are non-linear factor analysis (Bock et. al. 1988; McDonald, 1967) and DIMTEST (Stout et. al., 1991).

Unfortunately, these newer methods for appraising test dimensionality are designed to handle only dichotomously-scored data, which limits their applicability. This limitation is unfortunate because more and more educational tests comprise both items that are scored dichotomously, such as multiple-choice items, and items that are scored polytomously, such as

constructed-response items. Thus, appraising the dimensionality of current educational tests is not straightforward.

PCA and common factor analysis can be used to analyze the dimensionality of a test data that include both polytomous and dichotomous items; however, the mere presence of dichotomous/polytomous scoring differences among items may produce "meaningless" (artifactual) factors due the scoring differences among the item types. Furthermore, the expected measurement error of the items would not be consistent across item type. In particular, the error associated with a single multiple-choice item is expected to be high. As Dorans and Lawrence (1987) state "item level data is fraught with noise due to the unreliability of a single item, ... variation due to differences in item difficulty and examinee item responding strategies are likely to dominate item level analyses" (p. 84). For these reasons, many dimensionality analyses have used an item parceling (bundling) approach, where subsets of items, called parcels or bundles, are created from subsets of individual items (e.g., Cattell, 1956, Cattell & Burdsal, 1975; Dorans & Lawrence, 1987).

Bennett, Rock, and Wang (1991) provide a recent example of the use of bundling for evaluating the dimensionality of a test comprising both multiple-choice and constructed-response items: the Advanced Placement Computer Science Examination. This exam comprised 50 multiple-choice items (scored dichotomously) and 5 constructed-response items (scored on a ten-point scale). Bennett et al. grouped the multiple-choice items into five bundles of ten items each. Thus, each examinee had ten scores: five based on the five multiple-choice item bundles, and five based on the constructed-response items. The 10X10 correlation matrix of bundles and constructed response items was then factor-analyzed, rather than the original 55X55 matrix of

multiple-choice and constructed response items. There are numerous other applications of bundling in the evaluation of test dimensionality (e.g., Dorans & Lawrence, 1987; Lukhele & Sireci, 1994; Thissen, Wainer, & Wang, 1994). The typical bundling procedure followed in these studies is to form bundles within content area that are similar with respect to average item difficulty and variance.

Although bundling items reduces the problems associated with item-level data and item format scale differences, the effect of the bundling process itself on the detection of dimensionality is relatively unknown. A legitimate concern is that the creation of bundles may *obscure* unique and construct-relevant variance present in the item-level data. For example, suppose a 30-item test measures three unique dimensions. If two bundles of 15 items each were created, obviously, a dimensionality analysis could not recover the three dimensions. If three bundles of ten items each were created, the way in which the bundles were formed would greatly determine whether the three dimensions are recovered. Although this example is extreme and would not likely be followed in practice, it illustrates the more insidious problem: important dimensionality information may be lost when items are bundled.

The primary purpose of this paper is to assess the effect of item bundling on multidimensional test data. A secondary purpose is to compare three methods for assessing this dimensionality. The methods are compared using both simulated and real test data.

Method

Simulated Data

Eight multidimensional data sets were generated varying in terms of dimensionality, inter-dimensional correlation, and number of items loading on each dimension. The data were

simulated to have one-, two-, or three-dimensions. The interdimensional correlation matrix for the multidimensional data sets was specified to be either slightly correlated (.10) or highly correlated (.60) similar to Hambleton and Rovinelli (1986). Within each of the correlation conditions either equal or unequal number of items were specified to load on each factor. So, in the equal loading conditions for the two-dimensional data, the first 50 items defined the first factor while the second 50 items measured the second factor. In the unequal loading condition for the two-dimensional data, 67 items measured trait 1 while 33 items loaded on trait 2. For the three-dimensional data, under the equal loading condition, 34 items measured the first factor and 33 items were set to load on the second and third factors. For the three-dimensional data in the unequal loading condition, 50 items loaded on the first latent variable while 25 defined the remaining two factors.

Unidimensional data were also generated for comparative purposes. We wanted to ensure that our techniques would be able to detect unidimensional data once the items were bundled together. For the unidimensional data, all items loaded on one factor. Table 1 presents a summary of the simulation conditions for each data set. All simulation conditions were replicated 100 times.

For both the unidimensional and multidimensional data sets, 100 dichotomously scored items were simulated with the 3-parameter logistic model. One thousand examinees were drawn from a normal ability distribution (mean=0 and standard deviation=1). The items were specified such that b was uniformly distributed in the interval (-2, 2); a was uniformly distributed in the interval (0.6, 2.0); and c was uniformly distributed in the interval (0.00, 0.25) (Hambleton and Swaminathan, 1985; Hambleton and Rovinelli, 1986).

Bundling Procedure, Simulated Data

Two bundling techniques were used to create the item parcels. Since the size of the item bundle is a question of interest, each technique was used to create four different sized bundles. Bundles consisted of four, five, 10, or 20 items; therefore, analyses were performed on either 25, 20, 10, or five bundles of items. First, parallel item bundles were constructed using observed p-values (percentage of examinees correctly answering an item) thereby creating bundles with approximately equal mean difficulty and standard deviation.[1] To confirm that the bundles had nearly the same difficulty and standard deviation, one-way ANOVAs were performed between bundles within a subtest. Using α=.05, no significant differences were found between the bundles. These bundles are called difficulty bundles.

Next, items were grouped randomly. Here, 100 numbers were randomly generated from a uniform distribution. The items were then sorted according to the random numbers and parceled. These bundles are called random bundles. For both the difficulty and random bundling strategies, item bundle scores were computed for each examine and correlations were computed between the bundles. The dimensionality of all bundling strategies were analyzed separately

CPA DATA

The Uniform Certified Public Accountant (CPA) exam consists of four subtests: Auditing, Accounting and Reporting (ARE), Financial Accounting and Reporting (FARE), and Legal and Professional Responsibilities (LPR). In order to pass the CPA exam, examinees must pass all four sections of the test. The data come from two random samples of 4,032 examinees who took all four sections of CPA exam in November, 1996. Three types of items (multiple choice, other objectively scored answer format (OOAF), and essay) are used in the CPA exam.

OOAFs are questions that consist of, for example, 10 multiple true-false items. These items are usually locally dependent within a single OOAF question.

Bundling Procedure, CPA Data

For the Uniform CPA exam, two item bundling strategies were used. First, we inspected the dimensionality structure through the use of "item-type bundles". This bundling strategy allowed us to evaluate the effects of item type on dimensionality. These bundles comprised items that were of the same format (i.e. essay bundles, multiple-choice bundles, other objective answer format (OOAF) bundles). Bundles with 10-point scales were created since all essay and many OOAF questions are on this scale. One OOAF item was split into two bundles and two OOAF items were combined into one bundle in order to approximate this scale. The multiple-choice items were then bundled according to subtest and item difficulty. So, the multiple-choice items within a subtest, say Auditing, were ranked and then grouped according to their p-values using a method similar to that used to create the difficulty bundles in the simulated data. This technique ensured that the bundles had approximate equivalent difficulty distributions. Again, the total score of the multiple-choice item-type bundles were calculated using the appropriate weights when necessary.[2] This process resulted in 39 item-type bundles worth 10 points each.

Next, we measured dimensionality with "content bundles". These bundles comprised items that were listed in the same content area within a test section. This bundling strategy allowed us to evaluate the presence of dimensions linked directly to the content area measured on the exam. First, twenty content area bundles were constructed (using the test specifications) with point values ranging from 4 to 43. The total score of each bundles was calculated using the appropriate weights when necessary. In future analyses, content bundles might be created so that

they are equal in size.

We also assessed the dimensionality of the exam or subtest at the item level. At the item-level, our analyses were restricted to the objectively-scored dichotomous items. This method allowed us to assess the affect that content as well as item type has on the dimensionality of the exam.

Procedure

Three methods were used to assess the dimensionality of the data sets: principal components factor analysis (PCFA), maximum-likelihood factor analysis (MLFA), and multidimensional scaling (MDS). PCFA were performed on each of the item-bundle correlation matrices and on the inter-item tetrachoric correlations matrices. Dimensionality was assessed by inspection of plots of the eigenvalues. Typically, the number of significant factors is determined by the appearance of an "elbow" in the plot (Hambleton and Rovinelli 1986). The number of eigenvalues to the left of this elbow is usually taken to be the dimensionality. We also looked at the proportion of variance explained by the first eigenvalue. If a data set is unidimensional, then the first eigenvalue should explain a relatively large proportion of the variance. If it does not, the data may be multidimensional despite the appearance of one dominant eigenvalue. This could also signify a unidimensional data set with a great deal of error.

MLFA were performed on each of the item-bundle correlation matrices. MLFA allows us to test hypotheses regarding the number of factors that should be fit to the data using a likelihood ratio chi-square test. Here we tested the hypothesis of the fit of a one-factor model versus a model with enough factors to fit the data perfectly. A non-significant chi-square indicates good model-data fit. Of interest was the degree to which the chi-square would change

both across and within the simulation conditions. These analyses were only performed on the 10, 20, and 25 bundle data sets. In the 5 bundle condition, there simply were not enough variables to obtain useful results with the MLFA. At the item level, we encountered problems with non-gramian matrices.

The chi-square test is adversely affected by sample size and the large samples used with the CPA data rendered the chi-square test virtually useless. So, the Tucker-Lewis Rho coefficient was used to assess dimensionality of the CPA exam. This coefficient measures the adequacy of a model by expressing what percentage of the variance of a completely specified model is accounted for by a reduced model. If the coefficient is large, usually .9 or greater, the reduced model accounts for most of the correlation among the variables, thus any bias introduced by leaving out factors is negligibly small.

Promax rotation was used to aid in the interpretation of the factor structure of the CPA exam. This type of rotation was used, instead of varimax rotation, because it provides a more realistic model of the data. The underlying traits detected by the MLFA are most likely correlated, and promax rotation allows for this relationship among the latent traits. MLFA also provides a measure of the uniqueness of the variables. Uniqueness is the variance of a variable that is not explained by the retained factors.

The MDS analyses were conducted using the SPSS implementation of the ALSCAL algorithm with the Euclidean distance options. Dimensionality was determined by inspecting the fit values (STRESS and R^2) and interpretability of the various dimensional solutions. Typically STRESS values should be less than .10 and R^2 values should be greater than .90. If data are unidimensional, then the STRESS value might be as high as .15. Scree plots of STRESS

and R^2 values were also inspected to help determine the dimensionality of a data set. Like factor analysis, dimensionality is determined by an "elbow" in the plot; here, however, the elbow exactly marks the number of underlying dimensions. MDS was performed only in conditions where, at least, 10 variables could be analyzed; fewer variables yielded unstable estimates for the MDS scale values. So, MDS was only performed on the 20 bundle condition (5 items per bundle), 25 bundle condition (4 items per bundle), and on the items themselves.

Results

The results presented below are averages across 100 replications. Also, the results of the random and difficulty bundles were essentially identical for most of the analyses. The results of the difficulty bundles will only be presented unless the result of analyses from the random and difficulty bundles differed.

Unidimensional Data

PCFA: Results of the item bundle and item level factor analyses of the difficulty bundles for the unidimensional data sets are presented in Table 2. Unidimensionality was correctly identified for all item-bundle conditions. Not surprisingly, the proportion of variance accounted for increases as the number of bundles in a data set decreases. The dimensionality of the item-level condition is more difficult to determine. Examination of eigenvalue plots suggest either a one- or two dimensional solution; however, the much larger proportion of variance explained by the first factor, relative to the second factor, may lead some to conclude the data are unidimensional.

MLFA: Table 3 summarizes the chi-square statistic and degree of freedom for the MLFA. None of the chi-squares are statistically significant, indicating unidimensionality for all conditions. Also of note is that the ratio of the chi-square to its degrees of freedom decreases, albeit slightly,

as the bundle size increases indicating a better model-data fit in the 10 bundle condition than in the other two conditions.

MDS: Table 4 presents the STRESS and R^2 values for the MDS analyses of both the difficulty and random bundles. For the random bundles, MDS correctly identifies unidimensionality in the 20 bundle condition. The STRESS values in the 25 bundle condition as well as the item-level condition are slightly above the .15 criteria indicating that the data are multidimensional. However, R^2 values in addition to the scree plots of the STRESS values suggest unidimensional solutions. The MDS analyses failed to find a unidimensional solution for the difficulty bundles. The fit statistics for the difficulty bundles reveal poor fit in three dimensions (the most dimensions that could be specified and still yield stable results). The MDS analyses for both the 20 and 25 bundle condition suggest a multidimensional solution. This result is surprising given the fact that both the PCFA and MLFA found the same data to be unidimensional. Moreover, until this point the random bundle data performed in much the same way as the difficulty bundle data. In order to attempt to explain this apparent aberration, we examined the stimulus coordinates of one of the replications. We looked for relationships between the difficulty of the bundle, the variance of the bundle, and the stimulus coordinates. As expected, the p-values failed to be strongly correlated with stimulus coordinates. The variance of the bundles was strongly correlated with the stimulus coordinates; however, the same relationship was found between the stimulus coordinates of the random bundle data and the variance of the random bundles. The two remaining hypotheses for these surprising findings are that the MDS solutions are picking up on a subtle aspect of the difficulty bundling algorithm, or are having trouble fitting data with such little variation.

Two-dimensional data

PCFA: Table 5 summarizes the results of the PCFA of two dimensional data for all correlation and loading conditions. For the 5 and 10 bundle conditions across all correlation and loading conditions, the analyses incorrectly suggest unidimensionsality. From Table 5, it is clear that PCFA detects more dimensions as the number of bundles increases; uncovering the most latent variables when item-level data is used.

In the low correlation, equal loading (LCEL) and the low correlation, unequal loading (LCUL) conditions for the 20 bundle data, the eigenvalue plots suggest one, or possibly two, dimensions. In both of the conditions, examination of the percent of explained variance reveals one dominant factor and a weak second factor. In the LCEL and LCUL conditions for the 25 bundle data, the scree plots of eigenvalues in Table 5 correctly suggest a two dimensional solution while the proportion of variance explained again suggest one dominant factor and a weak second factor. At the item-level for both conditions, the eigenvalue plots revealed a two-dimensional solution; however, one could also argue that these plots showed a three- or four-dimensional solution. The proportion of variance explained showed a two factor solution.

In the high correlation, equal loading condition (HCEL), the scree plots of eigenvalues suggest a one or two-dimensional solution for the 20 and 25 bundle conditions. For both of these bundling conditions, the proportion of variance explained suggests only a unidimensional solution. The item-level analysis again shows multidimensionality. Here, the eigenvalue plot suggested a unidimensional solution but a three- or four-dimensional solution is not unreasonable. The proportion of variance revealed one dominant factor with a weak second factor. In the high correlation, unequal loading condition (HCUL), the scree plots show a one-

dimensional solution for the 20 bundle condition and a one- or two-dimensional solution for the 25 bundle condition. In both these conditions, the proportion of variance accounted for only suggest a unidimensional solution. For the item-level analysis, the eigenvalue plot again showed a unidimemsional solution but, again, a three- or four-dimensional solution could be interpreted from these plots. The explained variance suggests a two factor solution.

MLFA: Table 6 summarizes the chi-square statistics for the maximum-likelihood factor analysis. Across all correlation and loading conditions except HCUL, a three-dimensional solution was suggested by the chi-square statistic for the 10 bundle data; in the HCUL condition, a two dimensional solution was revealed. For both the LCEL and HCEL conditions, MLFA uncovered a four-dimensional solution for both the 20 and 25 bundle data sets. In the LCUL and HCUL, MLFA suggests a three-dimensional solution for the 20 bundle data and a four-dimensional solution for the 25 bundle data. Of interest here is the degree to which this ratio changes across and within data sets and correlation loading conditions. Within all correlation and loading conditions, the 25 bundle data appears to have best model-data fit when a one-factor model is fitted to the data. In both low correlation conditions, it continues to fit the model best when a two factor model is used.

MDS: Table 7 presents the results of the MDS analyses of the items and item bundles. Using our criteria of STRESS (near or lower than .1) and RSQ (above .9), unidimensionality is correctly rejected across all conditions. At the item-level, the STRESS and R^2 values suggest a three-dimensional solution across all correlation and loading conditions. However, we could also conclude a two-dimensional solution if we look at the scree plots of the STRESS and R^2 values. In the low correlation conditions, the fit statistic suggest a two-dimensional solution for the 20

and 25 bundle data. Good fit in one dimension was observed for the high correlation conditions for the 20 and 25 bundle data.

Three-Dimensional Data

PCFA: Table 8 summarizes the eigenvalues and proportion of variance explained across the 100 replications for the three-dimensional data. For all correlation conditions analyses of the 5 and 10 bundle data sets again led to a conclusion of unidimensionality. Within correlation conditions, we again find that the proportion of explained variance by the first factor decreases as the number of bundles increases. In both the LCUL and the LCEL conditions, eigenvalue plots showed the data to be unidimensional in both the 20 and 25 bundle data sets but choosing three dimensions for both of these data sets is not unreasonable. For both of these conditions, the percent of explained variance suggested dominant first factors with weak second and third factors. At the item level, scree plots showed the data to have at least 3 dimensions in the LCEL and LCUL conditions; however, an argument could be made that the LCEL condition has at least 6 factors while the LCUL condition has 5 or 6 factors. In the high correlation conditions, plots of the eigenvalues uncovered only one factor in the 20 bundle data but revealed one or possibly two dimensions in the 25 bundle data. When the proportion of variance was examined under these conditions, there appeared to be one dominant factor.

MLFA: Table 9 presents the chi-square statistics for the three-dimensional data. Inspection of the table clearly reveals that as the number of bundles decreases, the number of dimensions detected decreases. In the 10 bundle condition across all correlation and loading conditions, three dimensions are detected. In the 20 bundle condition across all correlation and loading conditions, four dimensions are detected. And, in the 25 bundle condition across all other

conditions, the chi-square statistic suggests a five dimensional solution is necessary to fit the data.

MDS: The STRESS and R-square fit statistics are summarized in Table 10 for the three-dimensional data. Here, the item-level analysis again reveals a three-dimensional solution is necessary to fit the data across all correlation and loading conditions. When the bundling conditions were examined, the fit statistics showed the data to be multidimensional. The fit statistics as well as the scree plots of the STRESS statistic suggest a two dimensional solution for both the 20 and 25 bundle data across all conditions. In most of these cases the STRESS values were not strictly at .10 but they were only slightly above (only .11) and the R^2 indicated acceptable fit at the two-dimensional level.

CPA Data

Table 11 reveals the results of the factor analysis on the Uniform CPA test data. The data appear unidimensional for both bundling strategies (item type and content area) as well as for the item-level analysis (this finding held across both random samples). Similar to the analysis of simulated data, the amount of variance explained increases as the bundle size increases. Here we see that, regardless of assessment strategy (items or bundles), the first factor is the dominant factor. In the item-level analysis, however, this factor accounts for only 12% of the variance in both samples; moreover, there are 6 eigenvalues greater than one. When item-type bundles are examined, the first factor accounts for nearly 37% of the variance across both samples. Finally, when content bundles are examined, the first factor accounts for almost 44% of the variance across both samples. Thus, at the item level, the exam appears to be multidimensional. This is probably due to a large amount error in the item-level data.

Table 12 summarizes the results from the MLFA. Here we see that when item-type bundles are analyzed, the three-factor solution accounts for about 95% of the variance in both samples. Very little is added by going from a three to a four-factor solution; only 1.5% more variance is explained by going from the three- to four-factor solution in Sample 1 and 1.4% in Sample 2. Similarly when content bundles are analyzed, the TL Rho criterion indicates a three-factor solution across both samples. Here, the three-factor solution accounts for nearly 97% of the variance in both samples.

Inspection of the factor loadings for both bundling strategies reveal that the factors appear to correspond to content area in both samples. The Auditing bundles appear to load strongly on the first factor. Interestingly, this factor also seems to pull out the LPR item-type bundles. On the second factor, there are strong factor loadings for the FARE and ARE bundles suggesting that these items may be measuring similar concepts. The third factor appears to correspond to the LPR and ARE content areas.

Table 13 shows the STRESS and R^2 values for the MDS analyses of the CPA exam. Analysis of the 20 content bundles suggested a one-dimensional. Analysis of the 39 item-type bundles suggested a two-dimensional solution. Figure 1 presents the scatterplot of the stimulus coordinates for the two-dimensional solution. In interpreting the two-dimensional solution for the item-type bundles, it appears that one dimension separated the bundles according to content (multiple-choice, OOAF, and essay bundles were created within content area). On this dimension, LPR items appear to separate from FARE items. The other dimension appears to be an item-type dimension separating essays from multiple choice items.

At the item-level, at least three dimensions underlie the data based on the STRESS and R^2 scores (see Table 13). The first dimension was moderately related to item difficulty (correlation between item p-values and the coordinates on this dimension was .80). The other dimensions were more difficult to interpret, but did seem to be related to some distinctions among the content areas.

Discussion

Item bundling is a popular method for handling the non-linearity problem of dichotomous items. While this method helps overcome some of the shortcomings of dichotomous data, its effect on dimensionality has been unclear. The results suggest that item bundles tend to obscure multidimensionality but analyses on the items themselves tend to overestimate dimensionality.

For the unidimensional data, virtually all item-level analyses overestimated dimensionality. When the data were bundled, all bundling conditions supported unidimensionality, with the exception of the MDS analysis on the "difficulty" bundles. However, in general, the results for the unidimensional data indicate that item bundling is useful for compensating the dimensionality- increasing effects of the error inherent in the item-level data.

For the multidimensional data as the number of bundles increased, so too did detection of dimensionality. It is hard to generalize beyond the specific conditions simulated, but for these data, it is clear that more than ten bundles are needed to identify the two- or three-dimensions underlying the data.

Across the different types of analyses the number of detected dimensions increased as the size of the bundle decreased. Analyses of the large item bundles (5 and 10 bundle data) using PCFA consistently detected unidimensionality regardless of the simulation condition. Using

PCFA on the smaller bundles in the low correlation conditions, we correctly assessed the dimensionality of the two and three dimensional data. However, in the high correlation condition, the multidimensionality of the data sets were harder to detect, especially in the unequal loading conditions. When PCFA was based on the items themselves, we overdetected the number of dimensions even in the unidimensional data.

Using MLFA, multidimensionality was detected for all bundling conditions. MLFA tended to overestimate the number of dimensions for all but the 10 bundle, three-dimensional data sets. MLFA correctly estimated the number of dimensions when the unidimensional data was analyzed. This suggest that MLFA may be a useful tool when one wants to detect the presence of multidimensionality in a data set.

Finally, using MDS, multidimensionality was detected. However, for the three-dimensional data, it was equivocal whether MDS supported a two- or three-dimensional solution. So, when the bundles were analyzed, MDS recovered two dimensions across all conditions. Also, MDS detected multidimensionality in the unidimensional data when the difficulty bundles were analyzed. These problems suggest that MDS may be adversely affected by the bundling procedure. When the items were analyzed, MDS recovered the correct number of dimensions in the multidimensional data sets across all conditions as well as the unidimensional data.

Our results definitely suggest that the proportion of bundles relative to the size of the data set is important in dimensionality analyses (i.e., the size and number of bundles is important). One relevant question becomes at what point do we find an optimal bundle size that reflects the underlying dimensionality of a data set without overestimating the dimensions (like the item-level data) or underestimating the dimensionality (like most of the bundle-level data)? For our

data sets, it appears that the use of smaller bundles might better capture the underlying dimensionality (i.e. more variables). This question can better be answered by examining more bundling conditions. One must also question if our larger bundles were unsuccessful because there were too few of them.

Our analyses concentrated on five bundling conditions (if you count the item-level analysis as one item per bundle) for a 100 item test. We chose to analyze a large test to reflect the conditions of the real data with which we were working. It might be useful to simulate a shorter test with smaller bundling conditions. It would be also interesting to see how our techniques detected dimensionality when smaller bundles are used.

A limitation of the current study is the method we used for bundling items. In practice, items are bundled within a content area. We bundled our data across content areas because of the way our multidimensional data was simulated. A more realistic situation might be for the items to load jointly on traits, say .60 on one and .40 on the other. If we were to do this, we could bundle within a content area and we could see if the bundles adequately capture the underlying construct. Because our bundles were created across traits, we could not tell if the bundles were loading on the correct dimensions. Nonetheless, this research does show that bundling may confound the dimensionality detection process. Further study is required to better understand the aspect of bundling that produces these effects.

References

Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement, 28,* 77-92.

Bock, R. D., Gibbons, R. D., & Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement, 12,* 261-280.

Cattell, R. B. (1956). Validation and intensification of the Sixteen Personality Factor Questionnaire. *Journal of Clinical Psychology, 12,* 205-214.

Cattell, R. B., & Burdsal, C. A., Jr. (1975). The radial parcel double factoring design: A solution to the item-vs.-parcel controversy. *Multivariate Behavioral Research, 10,* 165-179.

Chen, T. & Davison, M. L. (1996). A multidimensional scaling, paired comparisons approach to assessing unidimensionality in the Rasch model. In G. Engelhard & M. Wilson (Eds.) *Objective measurement: Theory into practice (Vol. 3)*. Norwood, NJ: Ablex.

De Ayala, R. J. & Hertzog, M. A. (1991). The assessment of dimensionality for use in item response theory. *Multivariate Behavioral Research, 26,* 765-792.

Dorans, N. J. & Lawrence, I. M. (1987). *The internal construct validity of the SAT.* Princeton, NJ: Educational Testing Services.

Green, S. B. (1983). Identifiability of spurious factors with linear factor analysis with binary items. *Applied Psychological Measurement, 7,* 3-13.

Hambleton, R. K. & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10,* 287-302.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9,* 139-164.

Hattie, J., Krakowski, K, Rogers, H. J., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. *Applied Psychological Measurement, 20,* 1-14.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.

Lukhele, R. & Sireci, S.G. (1994). *Combining the multiple-choice and essay portions of the GED Writing Skills Test onto an IRT scale using a priori weights. Technical Report No. 94-3,* GED Testing Service, American Council on Education, Washington, D.C.

McDonald, R. P. (1967). Non-linear factor analysis. *Psychometric Monograph* (No. 15).

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82-99.

Stout, W. F., Nandakumar, R., Junker, B., Chang, H. H., & Steidinger, D. (1991). DIMTEST and TESTSIM [Computer program]. Champaign: University of Illinois, Department of Statistics.

Thissen, D., Wainer, H., & Wang, X. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analysis of two tests. *Journal of Educational Measurement, 31*, 113-123.

Endnotes

[1] Here, items were sorted according to their p-values from hardest to easiest. They were then placed into, for example, 25 bundles of four items by assigning a number from one to 25 for each of the first 25 items. In this bundling condition, the items that ranked 26 to 50 in difficulty were assigned to bundles by reversing the order of the assignment numbers so that we now went down from 25 to one. The items ranking from 51 to 75 in difficulty were again placed in bundles by assigning numbers from one to 25. For the remaining 25 items, the assignment numbers were again reversed.

[2] In the CPA exam, some multiple choice items are differentially weighted according to content specifications. These weights were used prior to dimensionality analyses.

Table 1 Summary of Simulation Conditions for Each Data Set

			Number of Items		
Data Set	Traits	$r(\theta_1, \theta_2)$	First Trait	Second Trait	Third Trait
1	1	n/a	100	n/a	n/a
2	2	.10	50	50	0
3	2	.10	67	33	0
4	2	.60	50	50	0
5	2	.60	67	33	0
6	3	.10	34	33	33
7	3	.10	50	25	25
8	3	.60	34	33	33
9	3	.60	50	25	25

Table 2 Eigenvalues and Percent of Variance Accounted for Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor			
		I	II	III	IV
Unidimensional	5 Bundles				
	λ	4.36 (.03)	0.18 (.01)	0.17 (.01)	0.15 (.01)
	% of σ^2	87.26 (.70)	3.66 (.26)	3.31 (.23)	3.04 (.18)
	10 Bundles				
	λ	7.52 (.12)	0.36 (.03)	0.32 (.02)	0.31 (.02)
	% of σ^2	75.19 (1.16)	3.55 (.28)	3.25 (.19)	3.05 (.18)
	20 Bundles				
	λ	11.77 (.29)	0.62 (.03)	0.58 (.02)	0.55 (.02)
	% of σ^2	58.87 (1.44)	3.12 (.17)	2.89 (.12)	2.75 (.12)
	25 Bundles				
	λ	13.29 (.37)	0.72 (1.48)	0.67 (.03)	0.64 (.03)
	% of σ^2	53.18 (1.48)	2.87 (.13)	2.69 (.11)	2.57 (.10)
	Items				
	λ	40.28 (1.54)	4.88 (.51)	1.92 (.15)	1.73 (.08)
	% of σ^2	40.28 (1.54)	4.88 (.51)	1.92 (.15)	1.73 (.08)

Table 3 Maximum-Likelihood Factor Analysis Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor			
		I	II	III	IV
Unidimensional	10 Bundles				
	χ^2 (df)	36.96(35)	21.47 (26)	11.55 (18)	
	sd	8.53	5.56	3.92	
	20 Bundles				
	χ^2 (df)	180.46 (170)	145.49 (151)	117.91 (133)	94.94 (116)
	sd	18.16	14.58	12.30	10.42
	25 Bundles				
	χ^2 (df)	295.57 (275)	249.16 (251)	212.21 (228)	180.78 (206)
	sd	26.49	23.26	20.88	18.47

Table 4 STRESS and R-Square Values for MDS Solutions using Euclidean Distance

Dataset	# of Bundles	Dimension(s)					
		I	II	III	IV	V	VI
Unidimensional, Difficulty	20 Bundles						
	STRESS	0.40 (.04)	0.26 (.02)	0.17 (.01)			
	R^2	0.60 (.13)	0.75 (.08)	0.83 (.06)			
	25 Bundles						
	STRESS	0.44 (.04)	0.27 (.02)	0.20 (.01)			
	R^2	0.51 (.11)	0.69 (.06)	0.78 (.05)			
	Items						
	STRESS	0.17 (.02)	0.13 (.01)	0.11 (.01)	0.09 (.01)	0.08 (.01)	0.07 (.01)
	R^2	0.92 (.02)	0.95 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.01)	0.97 (.01)
Unidimensional, Random	20 Bundles						
	STRESS	0.15 (.03)	0.09 (.02)	0.06 (.01)			
	R^2	0.93 (.03)	0.97 (.01)	0.98 (.01)			
	25 Bundles						
	STRESS	0.16 (.03)	0.10 (.02)	0.07 (.01)			
	R^2	0.92 (.03)	0.96 (.02)	0.96 (.01)			
	Items						
	STRESS	0.17 (.02)	0.13 (.01)	0.11 (.01)	0.09 (.01)	0.08 (.01)	0.07 (.01)
	R^2	0.92 (.02)	0.95 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.01)	0.97 (.01)

Table 5 Eigenvalues and Percent of Variance Accounted for Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor I	II	III	IV
2-Dim (r=.10; 50/50)	5 Bundles				
	λ	3.76 (.09)	0.41 (.08)	0.31 (.02)	0.27 (.02)
	% of σ^2	75.19 (1.71)	8.26 (1.53)	6.13 (.43)	5.47 (.33)
	10 Bundles				
	λ	5.72 (.17)	0.88 (.16)	0.54 (.04)	0.49 (.03)
	% of σ^2	57.18 (1.72)	8.82 (1.56)	5.39 (.36)	4.90 (.28)
	20 Bundles				
	λ	7.79 (.30)	1.85 (.34)	0.84 (.06)	0.75 (.03)
	% of σ^2	38.95 (1.51)	9.24 (1.72)	4.19 (.28)	3.76 (.15)
	25 Bundles				
	λ	8.47 (.33)	2.24 (.39)	0.96 (.08)	0.84 (.03)
	% of σ^2	33.86 (.01)	8.95 (1.56)	3.86 (.31)	3.35 (.12)
	Items				
	λ	23.15 (1.10)	17.39 (.79)	3.18 (.27)	2.70 (.27)
	% of σ^2	23.15 (1.10)	17.39 (.79)	3.18 (.27)	2.70 (.27)
2-Dim (r=.60; 50/50)	5 Bundles				
	λ	4.12 (.05)	0.28 (.03)	0.23 (.01)	0.02 (.01)
	% of σ^2	82.38 (.98)	5.36 (.58)	4.54 (.23)	4.07 (.23)
	10 Bundles				
	λ	6.74 (.14)	0.53 (.06)	0.43 (.03)	0.39 (.02)
	% of σ^2	67.36 (1.36)	5.34 (.62)	4.29 (.28)	3.70 (.22)
	20 Bundles				
	λ	9.90 (.29)	1.02 (.12)	0.71 (.04)	0.67 (.03)
	% of σ^2	49.49 (1.45)	5.09 (.61)	3.57 (.18)	3.33 (.15)
	25 Bundles				
	λ	10.96 (.35)	1.26 (.17)	0.81 (.04)	0.76 (.03)
	% of σ^2	43.83 (1.41)	5.04 (.67)	3.25 (.15)	3.04 (.12)
	Items				
	λ	33.04 (1.55)	7.81 (.34)	3.55 (.35)	2.36 (.20)
	% of σ^2	33.04 (1.55)	7.81 (.34)	3.55 (.35)	2.36 (.20)
2-Dim (r=.10;67/33)	5 Bundles				
	λ	3.85 (.08)	0.38 (.07)	0.29 (.02)	0.25 (.02)
	% of σ^2	77.10 (1.65)	7.60 (1.41)	5.74 (.46)	5.06 (.35)
	10 Bundles				
	λ	5.99 (.18)	0.78 (.14)	0.52 (.03)	0.47 (.03)
	% of σ^2	59.92 (1.76)	7.79 (1.37)	5.16 (.33)	4.73 (.27)
	20 Bundles				
	λ	8.34 (.30)	1.58 (.27)	0.82 (.05)	0.74 (.03)
	% of σ^2	41.68 (1.52)	7.90 (1.38)	4.08 (.24)	3.70 (.15)
	25 Bundles				
	λ	9.09 (.35)	2.01 (.02)	0.92 (.06)	0.83 (.03)
	% of σ^2	36.35 (1.40)	8.03 (1.29)	3.68 (.23)	3.32 (.13)
	Items				
	λ	27.38 (1.39)	13.19 (.94)	3.67 (.36)	2.33 (.28)
	% of σ^2	27.38 (1.39)	13.19 (.94)	3.67 (.36)	2.33 (.28)
2-Dim (r=.60;67/33)	5 Bundles				
	λ	4.15 (.05)	0.26 (.03)	0.22 (.02)	0.20 (.01)
	% of σ^2	82.95 (1.05)	5.15 (.57)	4.40 (.30)	3.96 (.27)
	10 Bundles				
	λ	6.83 (.14)	0.51 (.06)	0.42 (.03)	0.38 (.02)
	% of σ^2	63.31 (1.40)	5.07 (.59)	4.20 (.26)	3.85 (.22)
	20 Bundles				
	λ	10.11 (.31)	0.97 (.12)	0.71 (.04)	0.65 (.03)
	% of σ^2	50.56 (1.53)	4.86 (.60)	3.53 (.19)	3.26 (.14)
	25 Bundles				
	λ	11.24 (.38)	1.16 (.14)	0.80 (.03)	0.75 (.03)
	% of σ^2	44.94 (.02)	4.63 (.57)	3.21 (.14)	3.00 (.10)
	Items				
	λ	33.96 (1.56)	6.92 (.37)	3.82 (.37)	2.15 (.19)
	% of σ^2	33.96 (1.56)	6.92 (.37)	3.82 (.37)	2.15 (.19)

Table 6 Chi-Square Statistic and Degrees of Freedom for Maximum-Likelihood Factor Analysis Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor				
		I	II	III	IV	V
2-Dim (r=.10; 50/50)	10 Bundles					
	χ^2 (df)	250.25 (35)	44.41 (26)	19.71 (18)		
	sd	146.72	15.54	6.99		
	20 Bundles					
	χ^2 (df)	1064.42 (170)	255.01 (151)	164.16 (133)	116.85 (116)	89.67 (100)
	sd	325.87	56.04	29.65	15.23	12.16
	25 Bundles					
	χ^2 (df)	1674.51 (275)	419.37 (251)	284.10 (228)	213.25 (206)	174.76 (185)
	sd	439.60	66.62	33.57	20.81	17.63
2-Dim (r=.60; 50/50)	10 Bundles					
	χ^2 (df)	123.49 (35)	39.64 (26)	16.63 (18)		
	sd	58.89	15.44	5.56		
	20 Bundles					
	χ^2 (df)	575.93 (170)	252.00 (151)	152.61 (133)	113.63 (116)	87.05 (100)
	sd	158.98	63.17	21.63	14.30	11.58
	25 Bundles					
	χ^2 (df)	874.37 (275)	415.96 (251)	269.65 (228)	209.46 (206)	172.68 (185)
	sd	179.90	63.46	29.57	20.69	18.39
2-Dim (r=.10;67/33)	10 Bundles					
	χ^2 (df)	232.73 (35)	44.61 (26)	18.63 (18)		
	sd	143.81	15.86	5.67		
	20 Bundles					
	χ^2 (df)	910.57 (170)	257.26 (151)	155.39 (133)	115.45 (116)	88.77 (100)
	sd	272.02	48.56	21.55	15.64	12.66
	25 Bundles					
	χ^2 (df)	1313.50 (275)	412.76 (251)	266.68 (228)	208.65 (206)	171.78 (185)
	sd	311.32	65.20	29.75	23.91	20.44
2-Dim (r=.60;67/33)	10 Bundles					
	χ^2 (df)	115.65 (35)	17.84 (18)	7.55 (11)		
	sd	58.83	5.12	3.02		
	20 Bundles					
	χ^2 (df)	519.89 (170)	257.08 (151)	152.79 (133)	116.37 (116)	89.89 (100)
	sd	139.43	56.38	20.03	14.67	12.99
	25 Bundles					
	χ^2 (df)	766.39 (275)	414.26 (251)	267.17 (228)	213.10 (206)	175.49 (185)
	sd	158.33	60.53	27.84	20.75	18.78

Table 7 STRESS and R-Square Values for MDS Solutions using Euclidean Distance, Difficulty

Dataset	# of Bundles	Dimension(s)					
		I	II	III	IV	V	VI
2-Dim (r=.10; 50/50)	20 Bundles						
	STRESS	0.20 (.04)	0.09 (.02)	0.06 (.01)			
	R^2	0.88 (.05)	0.97 (.01)	0.98 (.01)			
	25 Bundles						
	STRESS	0.23 (.04)	0.10 (.02)	0.07 (.01)			
	R^2	0.85 (.07)	0.96 (.02)	0.98 (.01)			
	Items						
	STRESS	0.24 (.02)	0.13 (.01)	0.10 (.01)	0.08 (.01)	0.07 (.00)	0.07 (.
	R^2	0.84 (.03)	0.93 (.01)	0.95 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.
2-Dim (r=.60; 50/50)	20 Bundles						
	STRESS	0.17 (.05)	0.09 (.02)	0.06 (.01)			
	R^2	0.91 (.07)	0.96 (.02)	0.98 (.01)			
	25 Bundles						
	STRESS	0.18 (.03)	0.10 (.02)	0.07 (.01)			
	R^2	0.91 (.04)	0.96 (02)	0.98 (.01)			
	Items						
	STRESS	0.19 (.02)	0.12 (.01)	0.10 (.01)	0.09 (.01)	0.08 (.01)	0.07 (.
	R^2	0.90 (.02)	0.94 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.01)	0.98 (.
2-Dim (r=.10;67/33)	20 Bundles						
	STRESS	0.21 (.05)	0.09 (.02)	0.06 (.01)			
	R^2	0.86 (.06)	0.96 (.02)	0.99 (.01)			
	25 Bundles						
	STRESS	0.21 (.04)	0.10 (.02)	0.07 (.01)			
	R^2	0.87 (.00)	0.96 (.01)	0.98 (.01)			
	Items						
	STRESS	0.24 (.02)	0.13 (.01)	0.10 (.01)	0.09 (.01)	0.08 (.00)	0.07 (.
	R^2	0.84 (.03)	0.93 (.011)	0.95 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.
2-Dim (r=.60;67/33)	20 Bundles						
	STRESS	0.17 (.03)	0.10 (.02)	0.06 (.02)			
	R^2	0.92 (.04)	0.96 (.02)	0.98 (.01)			
	25 Bundles						
	STRESS	0.18 (.03)	0.10 (.02)	0.07 (.01)			
	R^2	0.91 (.03)	0.96 (.02)	0.97 (.01)			
	Items						
	STRESS	0.19 (.02)	0.13 (.01)	0.10 (.01)	0.09 (.01)	0.08 (.01)	0.07 (.
	R^2	0.90 (.02)	0.94 (.01)	0.96 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.

Table 8 Eigenvalues and Percent of Variance Accounted for Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor			
		I	II	III	IV
3-Dim (r=.10; 34/33/33)	5 Bundles				
	λ	3.44 (.09)	0.50 (.07)	0.39 (.03)	0.35 (.02)
	% of σ^2	68.81 (1.89)	10.07 (1.30)	7.87 (.61)	6.97 (.45)
	10 Bundles				
	λ	4.95 (.18)	0.93 (.12)	0.71 (.07)	0.59 (.03)
	% of σ^2	49.51 (1.83)	9.30 (1.15)	7.11 (.68)	5.87 (.33)
	20 Bundles				
	λ	6.37 (.27)	1.72 (.20)	1.28 (.15)	0.86 (.03)
	% of σ^2	31.84(1.34)	8.61 (1.00)	6.40 (.76)	4.30 (.17)
	25 Bundles				
	λ	6.77 (.31)	2.09 (.25)	1.56 (.18)	0.96 (.04)
	% of σ^2	27.09 (1.24)	8.37 (1.00)	6.26 (.74)	3.85 (.17)
	Items				
	λ	15.61 (.80)	13.53 (.65)	11.54 (.58)	2.72 (.24)
	% of σ^2	15.61 (.80)	13.53 (.65)	11.54 (.58)	2.72 (.24)
3-Dim (r=.60; 34/33/33)	5 Bundles				
	λ	4.02 (.05)	0.29 (.03)	0.25 (.02)	0.23 (.02)
	% of σ^2	80.47 (1.10)	5.82 (.51)	5.04 (.34)	4.59 (.32)
	10 Bundles				
	λ	6.46 (.15)	0.55 (.05)	0.48 (.03)	0.43 (.02)
	% of σ^2	64.57 (1.50)	5.47 (.46)	4.75 (.27)	4.33 (.24)
	20 Bundles				
	λ	9.27 (.31)	0.99 (.10)	0.82 (.05)	0.72 (.03)
	% of σ^2	46.37 (1.56)	4.50 (.52)	4.10 (.27)	3.59 (.15)
	25 Bundles				
	λ	10.19 (.39)	1.20 (.10)	1.00 (.09)	0.81 (.03)
	% of σ^2	40.74 (1.56)	4.78 (.40)	4.00 (.31)	3.24 (.12)
	Items				
	λ	28.39 (1.34)	6.32 (.30)	5.47 (.27)	3.13(.28)
	% of σ^2	28.39 (1.34)	6.32 (.30)	5.47 (.27)	3.13(.28)
3-Dim (r=.10; 50/25/25)	5 Bundles				
	λ	3.52 (.09)	0.47 (.06)	0.28 (.03)	0.33 (.02)
	% of σ^2	70.44 (1.75)	9.48 (1.20)	7.67 (.63)	6.59 (.45)
	10 Bundles				
	λ	5.14 (.17)	0.88 (.11)	0.69 (.07)	0.58 (.03)
	% of σ^2	51.43 (1.71)	8.79 (1.15)	6.87 (.65)	5.76 (.32)
	20 Bundles				
	λ	6.75 (.30)	1.57 (.18)	1.22 (.14)	0.85 (.04)
	% of σ^2	33.76 (1.48)	7.87 (.89)	6.09 (.71)	4.27 (.19)
	25 Bundles				
	λ	7.22 (.31)	1.95 9.22)	1.50 (.19)	0.95 (.04)
	% of σ^2	28.87 (1.24)	7.79 (.88)	5.99 (.78)	3.78 (.17)
	Items				
	λ	20.12 (1.23)	11.17 (.67)	9.32 (.61)	3.07 (.29)
	% of σ^2	20.12 (1.23)	11.17 (.67)	9.32 (.61)	3.07 (.29)
3-Dim (r=.60; 50/25/25)	5 Bundles				
	λ	4.04 (.06)	0.28 (.02)	0.25 (.02)	0.22 (.02)
	% of σ^2	80.88 (1.17)	5.63 (.50)	4.98 (.34)	4.49 (.32)
	10 Bundles				
	λ	6.51 (.16)	0.54 (.05)	0.47 (.03)	0.43 (.02)
	% of σ^2	65.11 (1.61)	4.37 (.45)	4.67 (.26)	4.28 (.22)
	20 Bundles				
	λ	9.41 (.34)	0.96 (.09)	0.80 (.06)	0.71 (.03)
	% of σ^2	47.04 (.02)	4.80 (.47)	3.99 (.28)	3.55 (.15)
	25 Bundles				
	λ	10.35 (.41)	1.17 (.11)	0.94 (.07)	0.81 (.03)
	% of σ^2	41.39 (1.62)	4.67 (.43)	3.78 (.27)	3.22 (.13)
	Items				
	λ	29.20 (1.45)	6.34 (.29)	4.65 (.26)	3.28 (.27)
	% of σ^2	29.20 (1.45)	6.34 (.29)	4.65 (.26)	3.28 (.27)

Table 9 Chi-Square Statistic and Degrees of Freedom for Maximum-Likelihood Factor Analysis Using Simulated Test Data, Difficulty

Dataset	# of Bundles	Factor				
		I	II	III	IV	V
3-Dim (r=.10; 34/33/33)	10 Bundles					
	χ^2 (df)	257.19 (35)	77.70 (26)	23.29 (18)		
	sd	102.63	39.55	8.70		
	20 Bundles					
	χ^2 (df)	1101.08 (170)	495.97 (151)	193.38 (133)	135.74 (116)	99.90 (100)
	sd	291.70	134.85	34.49	21.64	16.79
	25 Bundles					
	χ^2 (df)	1709.04 (275)	836.35 (251)	333.87 (228)	249.35 (206)	196.81 (185)
	sd	447.31	233.12	50.68	36.12	28.91
3-Dim (r=.60; 34/33/33)	10 Bundles					
	χ^2 (df)	114.97 (35)	48.94 (26)	22.30 (18)		
	sd	40.34	17.83	7.67		
	20 Bundles					
	χ^2 (df)	541.55 (170)	306.72 (151)	190.24 (133)	131.08 (116)	98.55 (100)
	sd	118.71	60.32	31.68	18.34	14.24
	25 Bundles					
	χ^2 (df)	853.32 (275)	510.12 (251)	327.36 (228)	237.23 (206)	188.88 (185)
	sd	194.15	96.37	49.83	31.44	24.45
3-Dim (r=.10; 50/25/25)	10 Bundles					
	χ^2 (df)	247.50 (35)	77.60 (26)	23.34 (18)		
	sd	107.25	36.87	9.11		
	20 Bundles					
	χ^2 (df)	1022.38 (170)	453.32 (151)	193.87 (133)	131.19 (16)	97.63 (100)
	sd	272.42	128.51	33.80	19.11	13.78
	25 Bundles					
	χ^2 (df)	1482.10 (275)	741.04 (251)	339.43 (228)	242.62 (206)	193.50 (185)
	sd	302.13	165.18	51.36	25.22	19.86
3-Dim (r=.60; 50/25/25)	10 Bundles					
	χ^2 (df)	112.44 (35)	45.05 (26)	19.95 (18)		
	sd	42.91	18.20	7.32		
	20 Bundles					
	χ^2 (df)	522.09 (170)	299.03 (151)	191.37 (133)	130.97 (116)	98.68 (100)
	sd	151.67	72.50	38.21	25.50	19.76
	25 Bundles					
	χ^2 (df)	812.28 (275)	490.06 (251)	327.20 (228)	237.72 (206)	192.06 (185)
	sd	172.21	84.28	49.57	35.47	28.87

Table 10 STRESS and R-Square Values for MDS Solutions using Euclidean Distance, Difficulty

Dataset	# of Bundles	Dimension(s)					
		I	II	III	IV	V	VI
3-Dim (r=.10; 34/33/33)	20 Bundles						
	STRESS	0.21 (.05)	0.11 (.02)	0.06 (.01)			
	R^2	0.86 (.06)	0.95 (.03)	0.98 (.01)			
	25 Bundles						
	STRESS	0.22 (.04)	0.11 (.02)	0.07 (.01)			
	R^2	0.86 (.05)	0.94 (.02)	0.98 (.01)			
	Items						
	STRESS	0.23 (.02)	0.16 (.01)	0.10 (.01)	0.08 (.01)	0.07 (.00)	0.06 (.0
	R^2	0.85 (.03)	0.90 (.02)	0.94 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.0
3-Dim (r=.60; 34/33/33)	20 Bundles						
	STRESS	0.21 (.05)	0.11 (.02)	0.06 (.01)			
	R^2	0.86 (.07)	0.95 (.03)	0.98 (.01)			
	25 Bundles						
	STRESS	0.22 (.04)	0.11 (.02)	0.07 (.01)			
	R^2	0.86 (.05)	0.94 (.02)	0.98 (.01)			
	Items						
	STRESS	0.18 (.02)	0.13 (.01)	0.10 (.01)	0.08 (.01)	0.07 (.01)	0.07 (.0
	R^2	0.91 (.02)	0.94 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.01)	0.98 (.0
3-Dim (r=.10; 50/25/25)	20 Bundles						
	STRESS	0.21 (.05)	0.10 (.02)	0.06 (.01)			
	R^2	0.87 (.06)	0.95 (.02)	0.98 (.01)			
	25 Bundles						
	STRESS	0.22 (.04)	0.11 (.02)	0.07 (.01)			
	R^2	0.87 (.05)	0.95 (.02)	0.98 (.01)			
	Items						
	STRESS	0.22 (.01)	0.15 (.01)	0.10 (.01)	0.09 (.01)	0.07 (.00)	0.07 (.0
	R^2	0.86 (.03)	0.90 (.02)	0.94 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.0
3-Dim (r=.60; 50/25/25)	20 Bundles						
	STRESS	0.21 (.05)	0.11 (.02)	0.06 (.01)			
	R^2	0.86 (.07)	0.95 (.03)	0.98 (.01)			
	25 Bundles						
	STRESS	0.22 (.04)	0.11 (.02)	0.07 (.01)			
	R^2	0.86 (.05)	0.94 (.02)	0.98 (.01)			
	Items						
	STRESS	0.18 (.02)	0.13 (.01)	0.10 (.01)	0.09 (.01)	0.08 (.01)	0.07 (.0
	R^2	0.91 (.02)	0.94 (.01)	0.96 (.01)	0.97 (.01)	0.97 (.01)	0.98 (.0

Table 11 Eigenvalues and Percent of Variance Accounted for Using CPA Test Data

Assessment Strategy	Factor(s), Sample 1						Factor(s), Sample 2					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Item Bundles												
λ	14.21	2.02	1.39	1.10	.972	.933	14.30	1.99	1.36	1.12	.958	.906
% of σ^2	36.44	5.17	3.56	2.83	2.49	2.39	36.67	5.10	3.48	2.88	2.46	2.32
Content Bundles												
λ	8.79	1.34	1.04	.897	.793	.731	8.81	1.31	1.01	.883	.809	.761
% of σ^2	43.94	6.72	5.18	4.48	3.96	3.66	44.05	6.56	5.05	4.41	4.04	3.81
Items												
λ	11.94	2.73	1.89	1.39	1.21	1.02	11.82	3.01	1.62	1.42	1.29	1.10
% of σ^2	11.94	2.73	1.89	1.39	1.21	1.02	11.82	3.01	1.62	1.42	1.29	1.10

Table 12 Tucker-Lewis Rho Coefficients from MLFA of CPA Test Data

Assessment Strategy	Dimension(s), Sample 1				Dimension(s), Sample 2			
	I	II	III	IV	I	II	III	IV
Item Bundles								
TL Rho	.850	.922	.953	.968	.856	.925	.953	.967
Content Bundles								
TL Rho	.856	.932	.966	.982	.866	.937	.968	.982

Table 13 MDS Solution for CPA Test Data based on Euclidean Distances

Assessment Strategy	Dimension(s), Sample 1						Dimension(s), Sample 2					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Item Bundles												
STRESS	.194	.102	.079	.069	.056	.048	.185	.098	.075	.062	.053	.047
R^2	.928	.972	.980	.984	.988	.991	.935	.974	.983	.967	.990	.991
Content Bundles												
STRESS	.046	.028	.023				.044	.027	.020			
R^2	.995	.998	.998				.995	.998	.999			
Items*												
STRESS	.249	.189	.147	.121	.105	.094	.250	.189	.147	.123	.106	.095
R^2	.838	.874	.907	.927	.939	.947	.837	.876	.908	.926	.938	.946

*This is based on 100 randomly selected items. Due to limitations in the MDS program, all items could not be used.

Education Dept. 545-0262

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029878

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: "Effect of Item Bundling on the Assessment of Test Dimensionality"

Author(s): Karla L. Egan, Stephen G. Sireci, Hariharan Swaminathan, Kevin P. Sweeney

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Karla L. Egan	Printed Name/Position/Title: Karla L. Egan / Research Assistant
Organization/Address: University of Massachusetts Amherst, MA 01003	Telephone: 413/545-1539 FAX: 413/545-4181
	E-Mail Address: egan@soe.umass.edu Date: May 5, 1998

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com