Evaluating the Dimensionality of the 2001 PIRLS Reading Assessment: An application of DIMTEST with DESM and CFA

Minhee Seo, University of North Carolina at Greensboro, m_seo2@uncg.edu

Chia-Yi Chiu, Rutgers, The State University of New Jersey, chiu2@illinois.edu

Louis Roussos, Measured Progress, lroussos@measuredprogress.org

Abstract

The purpose of this study was to investigate the dimensionality of the reading assessment of 2001 Progress in International Reading Literacy Study (PIRLS) by DIMTEST with DIMTEST-based Effect Size Measure (DESM) and Confirmatory Factor Analysis (CFA). The 2001 PIRLS reading assessment was comprised of eight booklets, with accompanying items based on different reading passages. With respect to the potential problem of reading passage dependency on measurement precision and content-related validity in reading comprehension tests, the study examined the dimensionality of the reading assessments by booklet of reading passage; a total of ten pairs of booklets were tested. As implied in previous studies reporting unexpected high Type I error rate of DIMTEST (Froliech & Habing, 2003; Seo & Roussos, 2006; Stout, Froelich, & Gao, 2001), DIMTEST rejected the null for all the ten cases; A pair of booklets is not dimensionally similar. DESM used in tandem with DIMTEST resulted in six sets that were unidimensional, but the other four sets were not. Despite different underlying assumptions, several pairs were found to be unidimensional, which is common in DESM and CFA.

Keywords: *test dimensionality, DIMTEST, reading passage dependency, CFA, DESM*

Introduction

The verification of unidimensionality has been an issue of great interest in educational measurement for several decades due to the prevalent use of unidimensional Item Response Theory (IRT) in analyzing standardized tests (McDonald, 1981; Stout, 1987). Furthermore, many commonly used methods for analyzing standardized tests (for example, scoring, scaling, linking, equating, test assembly, and DIF analysis methods) assume the test data can be characterized by a single dominant dimension. Thus, before applying such procedures it is important to evaluate the appropriateness of the unidimensionality assumption (Hulin, Drasgow, & Parsons, 1983). Although the dimensionality analysis has historically received considerable attention, there is no general agreement among psychometricians as to which procedures are right to use (Hattie, Krakowski, Rogers, & Swaminathan, 1996; Stout, 1987).

Among various test procedures for testing unidimensionality, the DIMTEST T statistic has been known as one of the successful indicators of assessing unidimensionality given a theoretical justification (Elias, Hattie, & Douglas, 1998).

The Progress in International Reading Literacy Study (PIRLS) is a large international study of the reading literacy of young children around the world conducted by the International Association for the Evaluation of Educational Achievement (IEA). The reading assessment of 2001 PIRLS was administered to 4th graders in 35 countries. Each of those countries has their own uniqueness in terms of education and culture such as curriculum and educational level. With regard to the PIRLS reading assessment administered to the examinees from various countries and using the several booklets, each based on different reading passages, it is essential to ensure that all examinees are put on the common scale for the purpose of comparison. Therefore, testing unidimensionality of the PIRLS reading assessment might be a first step before scoring or applying further analysis.

Most reading comprehensive assessments are composed of several reading passages and accompanying items to measure examinees' reading proficiency. The subsets of items having corresponding passages are supposed to be unidimensional. That is, each examinee can be represented by a point on a common scale of reading for their reading proficiency. However, it has been reported that measuring reading proficiency has an affect by reading passage dependency in terms of its measurement precision and content-related validity in a reading comprehension assessment (Keenan & Betjemann, 2006). The passage dependency may lower the actual precision of the test score than the nominal precision determined by classical or IRT theory (Sereci, Thissen, & Wainer, 1991). Given the potential problem related to reading passage dependency, it might be necessary to evaluate dimensionality of the reading test by a subset of items with different reading topic and passage. Therefore, the present study investigated the dimensionality of the 2001 PIRLS reading assessment by booklet, each based on a different reading topic and passage using DIMTEST (Nandakumar & Stout, 1993; Stout, 1987; Stout, Froelich, & Gao, 2001) with a DIMTEST Effect Size Measure (DESM, Seo & Roussos, 2007) and Confirmatory Factor Analysis (CFA) using LISREL.

DIMTEST

DIMTEST is a widely studied method for testing the hypothesis of test unidimensionality as represented by local item independence. To apply the DIMTEST procedure, a dataset is divided by two subtests, the Assessment subtest (AT) and the Partitioning subtest (PT). One set of items that are supposed to be dimensionally homogenous is first selected for AT. The remaining set of items is grouped for PT. These items are used to partition examinees into score groups. Then, DIMTEST categorizes examinees according to their scores on PT and

calculates the covariances of pairs of AT items conditioned on the PT score. Based on the conditional covariances of AT items, DIMTEST produces a T statistic (normally distributed under the null hypothesis) and performs a hypothesis test. Following the principle of local independence, the conditional covariances of AT items assumed unidimensionality are zero. If the DIMTEST T statistic is close to zero, the null hypothesis is retained. It is concluded that a set of data is unidimensional. Even though DIMTEST performs well, the problem of unexpectedly high Type I error rates on DIMTEST has been noticed by several studies (Froliech & Habing, 2003; Seo & Roussos, 2006; Stout, Froelich, & Gao, 2001). As a matter of fact, DIMTEST can indicate statistical rejection even when the lack of unidimensionality is rather small even if the sample size is large enough to detect it.

DESM

Seo and Roussos (2007) have proposed DESM to improve the statistical performance of DIMTEST and to provide more information on the degree to which the test data depart from unidimensionality. The DESM procedure calculates the difference between the conditional covariance estimates of AT items on actual data and covariance estimates on unidimensional date. Then, the value of DESM indicates how far off the AT conditional covariances are from zero. The DESM statistic is given by

$$\left\{ \hat{E}_{\theta_{PT}} \left[\sum_{i<j \in AT} COV(X_i, X_j | \theta_{PT}) \right] - \hat{E}_{\theta_{PT}} \left[\sum_{i<j \in AT} COV(X_i, X_j | \theta_{PT}) | d = 1 \right] \right\} \times 100$$

where $COV(X_i, X_j | \theta_{PT})$: covariance between AT items i and j for examinees conditioning on true score on the PT subtest.

In a Monte Carlo study simulated in such various conditions like sample size and test length, Seo and Roussos (2007) found that DESM used in tandem with DIMTEST provides better statistical performance than DIMTEST used alone and DESM is insensitive to sample size. Based on the simulation study, they proposed a criterion for statistical testing unidimensionality using DIMTEST with DESM. DIMTEST null hypothesis that AT is dimensionally similar to PT is rejected only when DIMTEST T statistic is larger than the critical value and DESM is larger than .70 (originally reported at .20 on the DETECT scale). In a recent study having more various research conditions and evaluating the DESM estimator bias, Seo and Roussos (2008) suggested more conservative criterion for DESM. DIMTEST null hypothesis is rejected only when DIMTEST T statistic is larger than the critical value and DESM is larger than .60.

CFA

CFA is a commonly used method to test hypothesized factor model of data structure. CFA is often considered as a special case of structural equation modeling (Bollen, 1989). Despite the many fit indices developed for CFA, there has been little clear evidence for the choice of one or more appropriate indices. It has been well known that fit indices based on Chi-square statistic are sensitive to sample size (Bentler & Benett, 1980; Joreskog, 1978; McDonald, 1982). Marsh, Balla, and McDonald (1988) examined a large number of fit indices regarding the influence of sample size and found such five indices relatively independent of sample size. These include the Fitting Function Incremental type 2 index (FFI2), the Likelihood Ratio Incremental type 2 index (LHRI2), the Chi-square Incremental type 2 index (X^2I2), the Tucker-Lewis index (TLI, often referred to as Non-Normed Fit Index (NNFI)), and the Cudeck and Brown rescaled Akaike Information criterion incremental type 2 index (CAKI2). In later research investigating CFA fit indices in terms of the effect of sample size and model parsimony, Marsh and Balla (1994) suggested five more indices for goodness-of-fit statistics in CFA: McDonald's Mk index (also referred as RMSEA), McDonald's Mc index, the Goodness-of-Fit Index (GFI), the Adjusted Goodness-of-Fit Index (AGFI), and the Relative Noncentrality Index (RNI). According to Kline (1998), at least four for goodness-of-fit were recommended to use such as Chi-square; GFI, Normed Fit Index (NFI) or Comparative Fit Index (CFA); NNFI; and Standardized Root Mean square Residual (SRMR). With respect to the large sample size of data set for the study, NNFI, RMSEA, and GFI were chosen for fit index. The difference in Chi-squares between one and two factor models (Joreskog, 1978) was adopted for alternative approach.

Methodology

Data Source

For this empirical study, the data set of the 2001 PIRLS reading assessment was used (N= 146,490). The reading comprehension test was comprised of eight booklets with accompanying items, each booklet based on a different reading topic and passage. Each booklet had a set of 11 to 14 items. Two booklets were administered to each examinee. As a result, each booklet was exposed to 25% of the examinees. The characteristics of each booklet are listed in <Table 1>.

[Take in Table 1 about here]

Out of a total 98 items, 46 were multiple choice items. Seventy items were dichotomously scored. The pairs of booklets taken by the same examinee groups, the number of examinees,

and the number of items in each booklet studied are shown in <Table 2>.

[Take in Table 2 about here]

As listed in Table 2, there were ten separate groups according to the set of booklets they took. Unlike other groups which had 10,000 to 11,000 examinees, only one group, F, was composed of about 30,000 examinees.

Data Analysis Method

The dataset was analyzed using six methods: (a) DIMTEST used alone with All the valid cases (DIM-A); (b) DIMTEST used alone with Bootstrap method (DIM-B); (c) DIMTEST with DESM using All the valid cases (DESM-A); (d) DIMTEST with DESM using Bootstrap method (DESM-B); (e) CFA assuming one latent variable (CFA1), and (f) CFA assuming two latent variables by booklet (CFA2). Bootstrap method was adopted to compare the results of using all the cases to those of using a smaller sample. One of the factors causing Type I error inflation of DIMTST was a large sample size, especially when used for a short test. If it is found that the results of using bootstrap method are similar to those of using all the cases, it would be able to provide confidence to use a smaller sample rather than large data set. Furthermore, the current official version of DIMTEST allows users analyze a maximum of 12,000 examinees. Therefore, it might give a good example for those who need to analyze a big data set from large scale assessment. For analysis using the bootstrap method, a sample of 5000 examinees was randomly selected. According to Seo and Roussos (2008), tests using a sample size of 5000 provide a better result in terms of both Type I error rate and power.

Six methods were employed for each set of booklets separately. Examinees who did not reach all the items in a booklet were eliminated from the data analysis. Only items dichotomously scored were included. Because the reading assessment was designed to be unidimensional, DESM was expected to indicate unidimensionality or weak multidimensionality; that is a DESM below .60. For CFA, a model assuming one latent variable was expected to show a better fit than or not significantly different from a model assuming two latent variables by booklet.

Specifically, the data set was analyzed as follows.

DIM-A

1. A set of two booklets listed in Table 1 was selected.

2. The shorter booklet of the two was selected for AT and the other was assigned to PT.

3. A DIMTEST was run and the rejection for the DIMTEST statistic was recorded at the significant level of 0.05.

These three steps were done for all ten sets of booklet separately.

DIM-B

The analysis procedure for DIM-B was the same as the procedure of DIM-A, except for Step 3. For Step 3, a DIMTEST was run for 100 independent trials with resampling of 5000 examinees and the rejection rate was recorded at the significant level of 0.05.

DESM-A

DESM-A used the same procedure as DIM-A except that DESM-A ran both DIMTEST and DESM. The rejection was decided when DIMTEST T is larger the critical value at the significance level of .05 and DESM is larger than .60.

DESM-B

The analysis using DESM-B followed the same procedure of DESM-A, with resampling methods.

1. A set of two booklets listed in Table 1 was selected.

2. Shorter booklet of two was selected for AT and the other was assigned for PT.

3. DIMTEST and DESM were run for 100 independent trials with resampling.

4. The rejection rates were recorded when DIMTEST T is larger the critical value at the significance level of .05 and DESM is larger than .60.

CFA1

For CFA, LISREL 8.8 was employed. CFA1 assumed only one latent variable accounted for by all the items no matter what booklet they belong to. The latent variable was supposed to be reading ability which the 2001 PIRLS reading assessment is designed to measure. All the items in two booklets were prespecified to have direct relationship to a single latent variable, reading ability.

CFA2

CFA2 assumed two latent variables by reading passage dependency. The items in a booklet were assigned to measure the first latent variable supposed to be a specific dimension of reading ability by the booklet, while the items in the other booklet were prespecified to relate only to the second latent variable supposed to be additional dimension by the other booklet.

For CFA, the cutoff to consider a close fit of model for each index were adopted as RMSEA of .05 (Browne and Cudeck, 1993), and GFI and NNFI larger than .95 (Hu and Bentler, 1999; Schumacker and Lomax, 2004).

Findings and Discussion

The results using DIM-A and DIM-B were listed in Table 3.

[Take in Table 3 about here]

As implied in previous studies (Froliech & Habing, 2003; Seo & Roussos, 2006; Stout, Froelich, & Gao, 2001), the DIMTEST presented very high rejection rates. On ten sets, only two were decided to be unidimensional by DIM-A. These include: Booklet 1& Booklet 6 and Booklet 5 & Booklet 6. These two sets also showed lowest the rejection rates in DIM-B. The cases with a rejection rate below 50% in DIM-B were the same as the cases decided to be unidimensional by DIM-A. The results in Table 3 support that DIM-A and DIM-B provide consistent results.

As seen in Table 3, many sets of booklets in the 2001 PIRLS were observed as being non-unidimensional by DIMTEST used alone. The results in Table 4 by additional use of DESM showed a slight difference from the results in Table 3.

[Take in Table 4 about here]

Unlike the results in Table 3, six sets were concluded to be unidimensional by DESM-A: (a) Booklet 1 & Booklet 2; (b) Booklet 1 & Booklet 6; (c) Booklet 1 & Booklet 8; (d) Booklet 2 & Booklet 4; (e) Booklet 2 & Booklet 5; and (f) Booklet 5 & Booklet 6. Only two sets had the same results as DIM-A. The other four sets of booklets were decided not to be unidimensional. These include: (a) Booklet 3 & Booklet 7, Booklet 4 & Booklet 6, Booklet 4 & Booklet 8, and Booklet 5 & Booklet 8. As seen in Table 4, Booklet 1 was tested directly with Booklet 2, 6, and 8. All those three tests presented the same result of unidimensionality. Therefore, Booklet 1, 2, 6 and 8 might be considered as being unidimensional. Booklet 2 turned out to be unidimensional with Booklet 4 or Booklet 5. Considering that Booklet 2 was tested directly with one of them, those three booklets might be grouped together as being unidimensional. However, it was found that Booklet 4 was not dimensionally similar to Booklet 6 or Booklet 8.

The DESM-A results are shown in Figure 1.

[Take in Figure 1 about here]

DESM-B provided the rejection rates from 100 trials. The results of DESM-B were very similar to the DESM-A results. Six sets of booklets showing the rejection rate under 50% were those concluded to being unidimensional by DESM-B. The other four sets displayed relatively high rejection rates above 65%. The values of DESM by DESM-B were almost the same as those by DESM-A. The largest difference in the DESM values between the two methods was only .02. There were three sets of booklets showed the DESM values lower than .40; Booklet 1 & Booklet 2, Booklet 2 & Booklet 4, and Booklet 2 & Booklet 5. In the aspect that the value of DESM reflects the degree to which the test data depart from unidimensionality, those three sets could be considered as being unidimensional. The rejection rates for those three sets were, even, 0%. In contrast, two sets, (a) Booklet 4 & Booklet 6; and (b) Booklet 4 & Booklet 8 displayed a 100% rejection rate. These results were consistent over four methods; DIM-A, DIM-B, DESM-A, and DESM-B. The results of CFA were listed in Table 5.

[Take in Table 5 about here]

All the fit indices supported close fits of models in CFA no matter how many latent variables were assumed. However, the difference tests in Chi-square value were all statistically significant. The indices in CFA seemed not to be sensitive enough to distinguish two models with different model assumptions in terms of dimensionality by reading passage. With regard to the difference in fit indices between CFA1 and CFA2, two sets showed the smallest differences on all four indices. These include: Booklet 1 & Booklet 2, and Booklet 2 & Booklet 4. These sets might be considered as being relatively more unidimensional than other sets of booklets. However, there were several sets that displayed relatively large differences such as Booklet 3 & Booklet 7, Booklet 4 & Booklet 6, and Booklet 4 & Booklet 8. These results were consistent with the results of DESM-A and DESM-B.

Conclusion and Implications

With regard to the PIRLS reading assessment administered to the examinees from various countries and using several booklets, each based on different reading passages, it is important to test unidimensionality among sets of booklets in the reading assessment. Once unidimensionality of booklets was verified, the score for all examinees could be on the common scale for the purpose of comparison. Therefore, the present study evaluated the dimensionality of the 2001 PIRLS reading assessment by booklet. The evaluation was done through three procedures; DIMTEST, DESM used in tandem with DIMTEST, and CFA.

As noticed by previous studies (Froliech & Habing, 2003; Seo & Roussos, 2006; Stout, Froelich, & Gao, 2001), DIMTEST showed very high rejection rates. Of ten sets, DIMTEST resulted in eight sets that are not dimensionally similar. Only two sets of booklets were concluded to be unidimensional. However, somewhat different results were drawn when applying DESM with DIMTEST. As seen in Figure 1, Booklet 1, Booklet 2, Booklet 4, and Booklet 5 could be decided as being unidimensional according to the DESM-A and DESM-B results. While, Booklet 4 was found to be dimensionally different from either Booklet 6 or Booklet 8. However, it was noticed that fit indices in CFA is insensitive to distinguish dimensionality by modeling additional latent variable. Even though DIMTEST and CFA have different underlying assumptions, some common results were founded. For example, the set of Booklet 1 & Booklet 2 resulted being unidimensional and showed relatively small DESM while providing smallest difference in Chi-square, RMSEA, NNFI, and GFI between CFA1 and CFA2. Based on the results satisfying all four methods, it was concluded that Booklet 1, Booklet 2, and Booklet 4 were unidimensional, but Booklet 6 was dimensionally different from those booklets as shown in Figure 2.

[Take in Figure 2 about here]

DIM-A provided almost the same results as DIM-B. The results of DESM-A were also very similar to those of DESM-B. These results supported that resampling method performs well in dimensionality analysis. It would provide more confident to use resampling method, not use all the cases for those who deal with huge data set from large scale assessment. The findings of the study implied that DESM would be very useful when a unidimensionality statistical significant test such as DIMTEST is used with large sample sizes where rejection of the null hypothesis could occur for only minor violations of the assumption of unidimensionality as often seen in administering a large scale standardized test in practice. DESM would be also advantageous in multidimensional case by providing valuable information on the degree of multidimensionality and make the result more interpretable for non-unidimensional case.

References

Bentler, P. M., & Benett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

IEA. (2001). PIRLS 2001 user guide for the international database. Retrieved March, 2007, from http://pirls.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html

Elias, S., Hattie, J., & Douglas, G. (1998). *An assessment of various item response model and structural equation model fit indices to detect unidimensionality.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Froelich, A. & Habing, B. (2003). *Conditional covariance based subtest selection for DIMTEST.* Unpublished manuscript. Retrieved May 15, 2005, from http://www.public.iastate.edu/~amyf/ver021003.pdf

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.

Hu, L., & Bentler, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Hulin, C. L., Drasgow, F. & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement.* Homewood, IL: Dow & Jones-Irwin.

Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43,* 443-447.

Keenan, J. M. & Betjemann, R. S. (2006). Comprehending the gray oral reading test without reading it: Why comprehension test should not include passage-independent items. *Scientific Studies of Reading, 10*, 363-380.

Kline, R. B. (1999). *Principles and practice of structural equation modeling*. NY: Guilford Press.

Marsh, H., & Balla, J. R. (1994). Goodness-of-fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality and Quantity, 28*, 185-217.

Marsh, H., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

McDonald, R. P. (1981). The dimensionality of test and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6,* 379-396.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics, 18,* 41-68.

Schumacker, r. e., & Lomax, R. G. (2004). *A biginner's guide to structural equation modeling* (2nd Ed.). Manwah, NJ: Lawrence Erlbaum Associates.

Segar, A. (1997). Assessing the unidimensionality of measurement: a paradigm and illustration within the context of information systems. *Omega, 25*, 107-121.

Seo, M, & Roussos, L. (2006, March). *An Investigation of DIMTEST Type I error on short tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Seo, M., & Roussos, L. (2007, April). *Formulation of an effect size measure for DIMTEST.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Seo, M., & Roussos, L. (2008, July). *Formulation of a DESM and evaluation of DESM estimator bias.* Paper presented at the annual meeting of the Psychometric Society, Durham, NH.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*, 237-247.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrica, 52,* 589-617

Stout, W., Froelich, A., & Gao, F. (2001). Using resampling to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duijin, & T. A. B. Snijders (Eds), *Essays on item response theory* (pp.357-375). New York: Springer-Verlag.

Table 1: The Characteristics of Data by Booklet

Booklet	Topic	Required skill	Number of Items			
			Total	MC [a]	CR [b]	Dichotomous[c]
1	Antarctica	Acquire and use Information	11	4	7	7
2	Leonardo	Acquire and use Information	12	6	6	8
3	Pufflings	Acquire and use Information	13	8	5	10
4	River	Acquire and use Information	11	3	8	6
5	Clay	Literary Experience	13	6	7	10
6	Flower	Literary Experience	13	7	6	10
7	Hare	Literary Experience	11	5	6	7
8	Mice	Literary Experience	14	7	7	12
	Total	AI: 47, LE:51	98	46	52	70

[a] MC: Multiple choice items

[b] CR: Constructed response items

[c] Dichotomous: Dichotomously scored items including both multiple choice and constructed response items

Table 2: Availability of booklets taken by the same examinee group

Group ID (n valid) [a]	Booklet (number of items)[b]							
	1 (7)	2 (8)	3 (10)	4 (6)	5 (10)	6 (10)	7 (7)	8 (12)
A (10697)	X	X						
B (11090)	X					X		
C (11446)	X							X
D (10943)		X		X				
E (10787)		X			X			
F (29551)			X				X	
G (11343)				X		X		
H (11409)				X				X
I (10561)					X	X		
J (10875)					X			X

[a] n valid: the number of examinees after eliminating those who did not reach all the items in the corresponding booklet

[b] number of items: the number of dichotomously scored items in the corresponding booklet

Table 3: Results of DIMTEST used alone

Booklet Set	DIM-A[a]		DIM-B[b]
	Reject null	results	Rejection Rate
1 & 2	Yes	Non-unidimensional	66%
1 & 6	No	Unidimensional	41%
1 & 8	Yes	Non-unidimensional	100%
2& 4	Yes	Non-unidimensional	68%
2& 5	Yes	Non-unidimensional	100%
3& 7	Yes	Non-unidimensional	90%
4& 6	Yes	Non-unidimensional	100%
4& 8	Yes	Non-unidimensional	100%
5& 6	No	Unidimensional	50%
5& 8	Yes	Non-unidimensional	79%

[a] DIM-A: DIMTEST used alone on all the valid cases

[b] DIM-B: DIMTEST used alone by bootstrap method (100 trials)

Table 4: Results of DESM used in tandem with DIMTEST

Booklet Set	DESM-A[a]		DESM-B[b]	
	Reject null	DESM	Rejection Rate	DESM
1 & 2	No (unidimensional)	.38	0%	.37
1 & 6	No (unidimensional)	.63	27%	.62
1 & 8	No (unidimensional)	.56	18%	.56
2& 4	No (unidimensional)	.20	0%	.19
2& 5	No (unidimensional)	.30	0%	.30
3& 7	Yes (not unidimensional)	.63	65%	.64
4& 6	Yes (not unidimensional)	.91	100%	.91
4& 8	Yes (not unidimensional)	.95	100%	.97
5& 6	No (unidimensional)	.55	10%	.56
5& 8	Yes (not unidimensional)	.70	77%	.70

[a] DESM-A: DESM used in tandem with DIMTEST on all the valid cases

[b] DESM-B: DESM used in tandem with DIMTEST by bootstrap method (100 trials)

Table 5: Fit indices of CFA using LISREL

Set of Booklets	Results of CFA1 [a] (Results of CFA2 [a])				
	X^2	Diff X^2 (p)	RMSEA	NNFI	GFI
1 & 2	743.3 (707.5)	35.8 (>.99)	0.026 (0.025)	0.98 (0.98)	0.99 (0.99)
1 & 6	983.5 (683.2)	300.3 (>.99)	0.026 (0.021)	0.99 (0.99)	0.99 (0.99)
1 & 8	1489.1 (1341.7)	147.4 (>.99)	0.032 (0.030)	0.98 (0.98)	0.98 (0.99)
2& 4	695.9 (611.1)	84.8 (>.99)	0.027 (0.025)	0.98 (0.98)	0.99 (0.99)
2& 5	1036.5 (885.5)	151.1 (>.99)	0.025 (0.023)	0.98 (0.99)	0.99 (0.99)
3& 7	5784.4 (3359.4)	2425.1 (>.99)	0.040 (0.030)	0.95 (0.97)	0.98 (0.99)
4& 6	1163.4 (672.0)	491.4 (>.99)	0.030 (0.022)	0.98 (0.99)	0.99 (0.99)
4& 8	1774.7 (1285.3)	489.4 (>.99)	0.033 (0.027)	0.98 (0.99)	0.98 (0.99)
5& 6	1281.3 (833.5)	447.8 (>.99)	0.025 (0.019)	0.99 (0.99)	0.99 (0.99)
5& 8	2323.1 (1608.0)	715.1 (>.99)	0.030 (0.025)	0.98 (0.99)	0.98 (0.99)

[a] CFA1: CFA with one latent variable

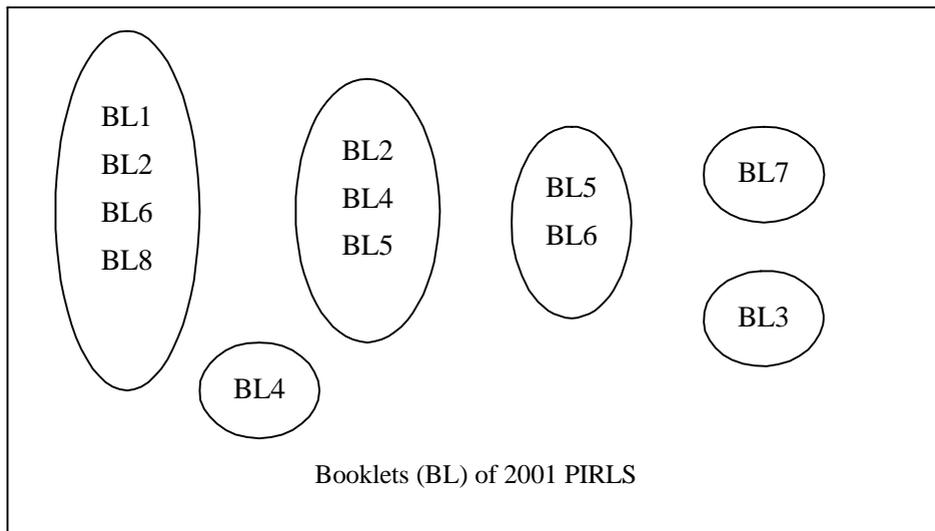[b] CFA2: CFA with two latent variables by booklet

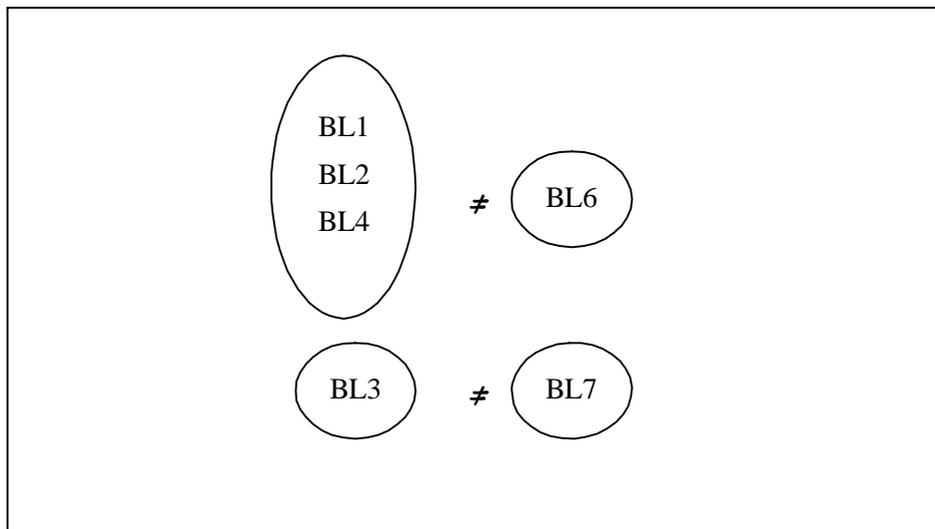Fig.1: Grouping of Booklets Unidimensional by DIMTEST with DESM

Fig.2: Dimensionality of Booklets Commonly by DIMTEST with DESM and CFA