A Modified Nonparametric Dimensionality Assessment (DIMTEST) for Short Tests

EunYoung Lim

Louis Roussos

Department of Educational Psychology
University of Illinois at Urbana-Champaign

The purpose of this research is to evaluate a modification of the DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001) dimensionality assessment method to make it more amenable to short tests, using the newest version of DIMTEST. This alternative DIMTEST uses item response vectors instead of sum scores to partition the examinees, and this method might overcome the disadvantage of not having enough partitioning scores with short tests, which is suspected to be the cause of inflated DIMTEST type 1 error results with short tests.

DIMTEST is based on item pair conditional covariance. Since DIMTEST estimates covariances conditioned on examinees' latent ability, DIMTEST requires the estimation of their latent abilities. The present version of DIMTEST employs the number correct score on an appropriately chosen subtest as an estimate of examinees' latent ability level. The number correct score depends on the number of items on the subtest, and the length of the subtest must result in sufficient separation of the examinees into different ability groups so as to yield adequate Type 1 error rate.

DIMTEST tests the null hypothesis of the unidimensional structure of a test. In order to test the null hypothesis of the unidimensional structure of a test, an Assessment

Subtest (AT) and Partitioning Subtest (PT) are required: conditional covariances of all pairs of AT items given PT item scores are estimated. If the AT items are dimensionaaly simiar to PT items, the conditional covariances of the AT items would be zero. In DIMTEST, there are some recommendations for the number of items for the Assessment Subtest (AT) and Partitioning Subtest (PT). The recommended minimum number of items for the PT is at least 15 and that of the AT is at least 3 (Stout et al, 1996). That is, the minimum total number of items on a test should be at least 18. Thus, we need a new approach for a short test, which has fewer than 18 items, in order to use nonparametric dimensionality assessments with appropriate power and Type I error.

In this study, we evaluated the effectiveness of using the item response vectors of the examinees on PT as an estimate of the examinees' ability level, instead of the number correct scores, for tests having fewer than 20 items. The PT item response vector is the examinees' response pattern on PT, with right answers scored as 1 and wrong answers scored as 0. For example, if an examinee has right answers on the first, third and fifth questions on a five-item test, the item response vector for the examinee is (1, 0, 1, 0, 1).

Theoretical background

Since the unidimensionality of a test is one of the most common assumptions made in using Item Response Theory, many researchers have studied how to assess the dimensionality structure of a test. Over the course of three decades, many methods have been developed, such as linear factor analysis, non-linear factor analysis, and nonparametric methods (Nandakumar et al, 1998).

The nonparametric methods, DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001), HCA/CCPROX (Roussos, Stout & Marden 1998) and DETECT (Zang & Stout, 1999), have been popular for detecting the dimensionality structure, since they avoid particular problems associated with parametric dimensionality analyses. First, nonparametric dimensionality assessments only assume the monotonicity the item response function (IRF), whereas parametric methods, such as nonlinear factor analysis, have a strong parametric modeling assumption for the IRF. Also, nonparametric methods are computationally simpler than parametric methods. Thus, nonparametric methods enable more efficient data analysis for detecting multidimensionality (Roussos, Stout, & Marden, 1998).

Item Pair Conditional Covariance

The DIMTEST nonparametric dimensionality method is based on item pair conditional covariances. Item pair conditional covariance is that covariance between two different items of a test conditioned on the examinees' latent ability. If these pairwise conditional covariances are aproximately zero, the result is pairwise local independence also referred to as WLI (weak local independence.) Strong local independence (SLI) means that each item, given examinees' latent traits, is independent from all other possible combinations of the items, and SLI is what is theoretically required to determine dimensionality. If strong local independence is held for all items of the test, WLI will also hold. Therefore, testing or measuring the conditional covariance for all pair items, given appropriate selected unidimensional examinees' latent ability, can be used an approximate manner to test for unidimensionality of a test.

The number correct test score is used for appropriate selected unidimensional examinees' latent ability for DIMTEST. Since examinee score vector is a sufficient statistic for IRT ability, examinees' item response vectors might be used for examinees' latent ability as a substitute for the number correct score.

Method and/or Technique

The item response vector is the examinees' answer patterns on a test. If there are 10 items on a test, each examinee has one of 11 total scores from 0 to 10. However, the number of item response vectors with 10 items is 2^{10} (1024). Thus, after selecting the AT items, the item response vectors for the PT items can be used to partition examinees. For example, if the number of AT items is 3 out of 10, the number of possible scores on the PT is 0. However, the number of item response vectors is 2^{7} (128). Therefore, even when a test has a small number of items, the limitation on the number of remaining items on a short test can be overcome by using the item response vectors of the examinees.

However, there is a high possibility of zero examinees or only one examinee for a certain item response vector. The minimum number of examinees for a PT cell is two examinees. Examinee cells with 0 or 1 examinees cells will be eliminated during the DIMTEST process. In this case, any item response vectors with 0 or 1 examinees will be excluded from analysis. Thus, we may need to merge similar item response vectors to prevent too great a proportion of the sample from falling into these sparse cells. In order to judge whether item response vectors are similar enough to merge together, we employed the Hierarchical Agglomerative Clustering (HAC) computer program. HAC was created by Roussos (1992) to cluster objects based on a pairwise proximity matrix. In this research, the objects to be clustered were examinees based on the pairwise priximity values of their PT item response vectors. Moreover, in order to estimate the pairwise proximity of item responses vectors, we calculated the Hamming distance and the P-weighted Hamming distance among all examinees.

DIMTEST

The newest version of DIMTEST (Stout, Froelich, & Gao, 2001) has overcome the limitation of selecting items for the Assessment Subtest (AT). The old version of DIMTEST (Stout, 1987) required two AT's (AT1 and AT2) and a PT (Partitioning Subtest.) AT2 was necessary to correct the positive bias that occurs when using only AT1 (Stout, 1987). However, since the AT2 items were selected from the remaining items after selecting the AT1 items, this severely limited the use of DIMTEST on short tests (e.g. less than 20 items). Even though the new version of DIMTEST has improved upon this disadvantage of the old version, there is still the possibility of the same limitation with very short tests (e.g., less than 15 items).

The newest version of DIMTEST employs kernel smoothing to estimate the unidimensional IRFs. Based on the unidimensional IRFs, new sets of data are generated, and DIMTEST statistics are estimated. The DIMTEST statistic on the unidimensional data is an estimate of the positive bias that occurs in the case of unidimensionality. Then the DIMTEST statistic from the generated data is subtracted from the DIMTEST statistic from the real data to estimate the final DIMTEST statistic. For a more precise description, see Stout, Froelich, & Gao, (2001).

Estimating proximity of item response vectors

Hamming distance is the distance between two equal length vectors. Hamming distance is the sum of the absolute distances between two elements of two vectors in the same location. Equation (1) is a hamming distance. For example, there are two vectors (1,0,0,1,1) and (1,1,0,0,1). The hamming distance between these two vectors is 2.

$$|1-1|+|0-1|+|0-0|+|1-0|+|1-1|=2 \qquad (0)$$

In this research, two hamming distances were employed as substitutes to calculate the pairwise proximity values of item response vectors: hamming distance (1) and P-weighted hamming distance (2). P-weighted hamming distance weights each of the above difference in equation (0) by the proportion-right score of each item.

$$\text{Hamming distance} = \sum_{i=1}^{N} |u(i, j_1) - u(i, j_2)| \qquad (1)$$

$$\text{P-weighted Hamming distance} = \sum_{i=1}^{N} |u(i, j_1) - u(i, j_2)| P_i \qquad (2)$$

i=item, *j*=examinee, N=number of items, P_i is proportion-right score of item *i*, and $U(i, j)$ is the response of examine *j* to item *i* (dichotomously scored as either 0 or 1).

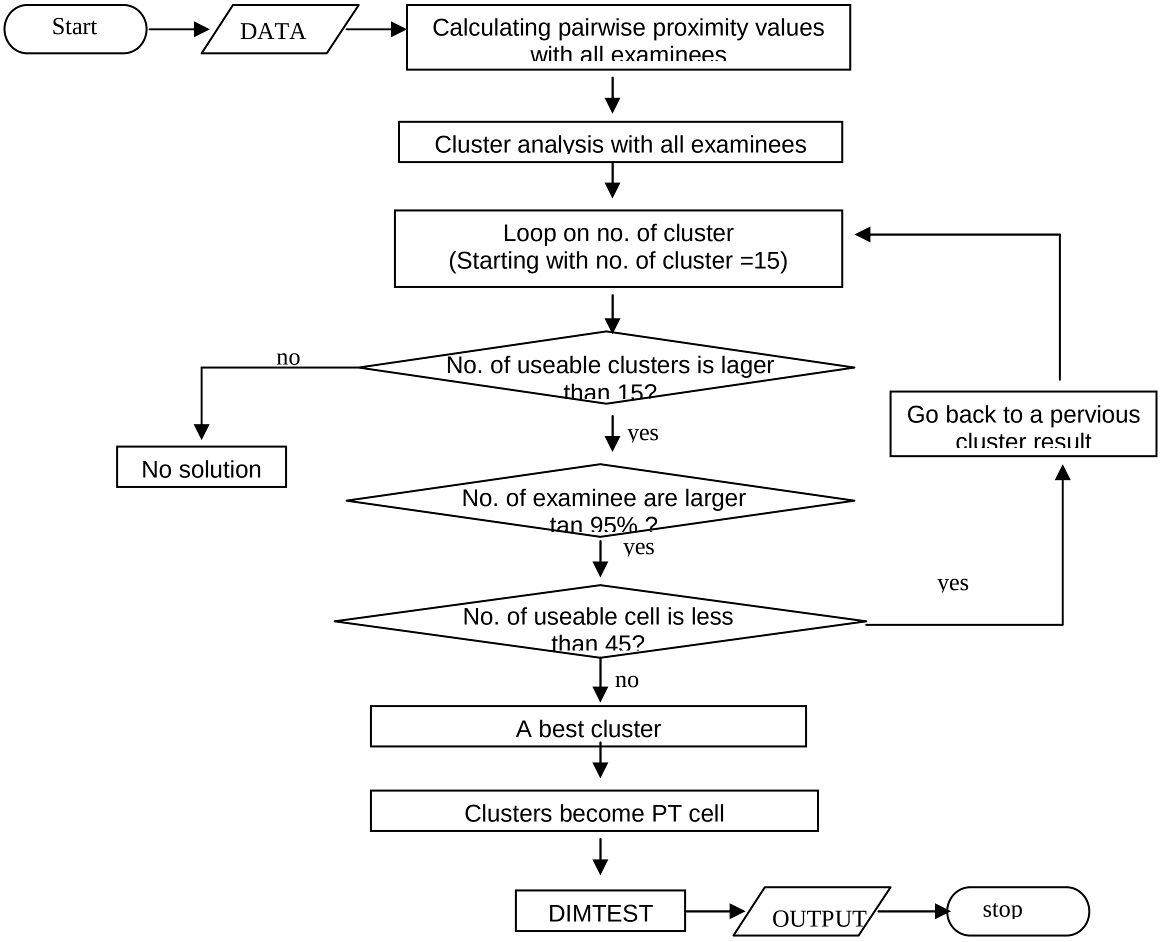
HAC

HAC has been used with CCPROX for clustering items which measure the same dimension. If pairwise items measure the same dimension, these two items will be closer than the two items in different dimensions. Based on this assumption, similar items belong to the same cluster (Roussos, 1992; Roussos, Stout, and Marden, 1998). In this research, we use HAC with our new proximity measures to obtain clusters that will be used as PT cells. The unweighted pair-group method of average (UPGMA) is used for the cluster method. HAC starts with each examinee on each cluster. Based on pairwise proximity values of examinees, examinees continuously merge one cluster with another cluster until reaching the final cluster level that includes all examinees in one cluster. Thus, if the data has 100 examinees, there are 100 cluster levels.

After finishing the whole cluster analysis, we find the appropriate number of clusters by working backward from the final cluster level. Since the sample size of our simulation goes up to 1000, and the number of possible PT item response vectors might become remarkably large, we made some rules for clustering examinees. First, there are upper and lower limits on the number of clusters. The lower limit is 15, since the recommended number for PT cells is at least 15. Thus, the cluster level with 15 clusters is the starting level for the backward method. The upper limit is 45, since tests with 40 PT cells empirically have provided appropriate DIMTEST results. Furthermore, if we have too many PT cells, there could end up being too many sparse cells for suitable DIMTEST calculations. For the appropriate appropriate number of clusters, the program starts looping to find whether the number of usable clusters is larger than 15, and checking the number of examinees. Usable clusters will be used as PT cells. Since DIMTEST's PT cells should include at least 2 examinees, we need to find usable clusters that have at least two examinees. Thus, 15 usable cells are required for the minimum number of PT cells, rather than 15 clusters. When 15 usable clusters are confirmed, the number of examinees are checked. If the number of usable clusters is 15 and over 95% of the examinees are included in the 15 usable clusters, the program continues back to the previous cluster level (which has one more cluster) to check the number of usable clusters and the number of examinees, until the usable clusters include 95% of the examinees.

Also, we have a special case: if the number of usable clusters is equal to 15, 87% of the examinees are allowed. Figure 1 shows the general process of DIMTEST using item response vectors as PT cells.

Figure 1 a procedure of DIMTEST with item response vectors



Simulation study

For this study, first we evaluated the effectiveness of DIMTEST for five short tests of different lengths (7 items, 8 items, 10 items, 15 items, and 18 items), using the number correct score on the PT as the examinees’ trait level. Second, we evaluated the effectiveness of DIMTEST for the short tests (7 items, 8items, 10 items, 15 items, and 18 items), using the item response vectors of the examinees.

In order to perform this research, a Type I error study and a power study were conducted with simulated data, and the number of unidimensionality rejections out of 100 trials were counted for various sample sizes (N=1000, N=1500, etc). Also, this research uses both unidimensional data and multidimensional simulation data. Each examinee for the unidimensional data will be randomly generated from standard normal distribution, and each examinee for the multidimensional data will be generated from standard bivariate normal distribution. Also, in order to generate data, we used the 3 PL model for item parameters. For a parameter, we used *log a* which follows a normal distribution with mean 0 and standard deviation 0.35. Also, *b* parameter follow a normal distribution with mean 0 and SD 1, and the lower and upper bound is .25 to 2.5. The *c* parameter is fixed as 0.15.

Results

We evaluated the effectiveness of DIMTEST for five short tests of different lengths (7 items, 8 items, 10 items, 15 items, and 18 items). The newest version of DIMTEST and two alternatives of DIMTEST were evaluated and compared. Since the newest version of DIMTEST (DIMTEST v.2.0.) excludes examinees that have all wrong answers and all right answers, one more alternative DIMTEST (DIMTEST v.2.1) that includes all examinees, regardless of their total scores, was employed. However, there is the possibility that examinees are excluded from this alternative DIMTEST analysis, since it is possible for an examinee to be the only one for a particular PT cell which would not be used in the DIMTEST. The other alternative DIMTEST is our new approach using the cluster method with a proximity matrix based on the item response vectors of the examinees.

DIMTEST v.2.0 with short tests

In order to evaluate the newest version of DIMTEST (DIMTEST v.2.0) with short tests, three different numbers of items (10, 15, and 18) and two and three different numbers of AT and PT sets were used for 100 trials. Table 1 shows the number of rejections out of 100 trials with unidimensional simulated data.

Table 1. Number of rejections out of 100 trials with 1 dimension (Type I error)

		No. of items (AT/PT)					
		10	15		18		
		(4/6)	(7/8)	(4/11)	(4/14)	(6/12)	(8/10)
sample size	1000	12	6	1	3	5	7
	2000	24	14	7	13	13	16
	4000	30	28	12	6	20	25

As the number of PT items increased, overall the rejection rate decreased. For example, when the number of PT items is 14, the rejection rate is 0.03 with a rejection level of 0.05. The suggested number of PT items is at least 15 (Stout et al, 1996). So it is not surprising that higher Type 1 error rates occur as sample size increases since all of the PT's here are less than 15.

Table 2. Number of rejections out of 100 trials with 2 dimensions (Power test)

		No. of items (AT/PT)					
		10	15		18		
		(4/6)	(7/8)	(4/11)	(4/14)	(6/12)	(8/10)
sample size	1000	95	99	99	97	99	98
	2000	100	100	100	96	100	100
	4000	100	100	100	100	100	100

Table 2 shows the number of rejections of 100 trials with 2 dimensions. DIMTEST with short tests is very powerful, but power without good Type 1 error rates must always be viewed cautiously.

Table 3 shows the proportion of the number of used examinees for DIMTEST. Tests with the smallest number of PT items use around 85% of the examinees. The small number of PT items has fewer PT sets, and the greater proportion of the examinees fall into the PT cells corresponding to all. Thus, we next evaluated whether increasing the number of used examinees can improve DIMTEST statistics, especially on shorter tests.

Table 3 proportion of examinees used

			No. of items (AT/PT)					
			10	15		18		
			(4/6)	(7/8)	(4/11)	(4/14)	(6/12)	(8/10)
type 1 error	1000	Mean	0.852	0.907	0.938	0.959	0.953	0.937
		SD	0.059	0.038	0.039	0.019	0.026	0.032
	2000	Mean	0.857	0.909	0.942	0.962	0.949	0.934
		SD	0.047	0.036	0.04	0.02	0.026	0.036
	4000	Mean	0.852	0.91	0.946	0.964	0.952	0.938
		SD	0.067	0.036	0.024	0.018	0.024	0.025
power	1000	Mean	0.843	0.911	0.944	0.967	0.956	0.944
		SD	0.069	0.039	0.03	0.017	0.022	0.03
	2000	Mean	0.845	0.911	0.949	0.967	0.953	0.938
		SD	0.056	0.038	0.029	0.018	0.022	0.029
	4000	Mean	0.844	0.912	0.945	0.967	0.953	0.939
		SD	0.06	0.037	0.026	0.018	0.022	0.028

DIMTEST v.2.0 vs. DIMTEST v.2.1

Table 4 shows the proportion of used examinees used in the Type 1 error study and power study. The range of the proportion of examinees is 0.725-0.9675 with DIMTEST v.2.0, and the range of the proportion of examinees is 0.953-1.00 with

DIMTEST v.2.1 including all examinees. DIMTES v.2.1 clearly improved on proportion of examinees used in comparison to DIMTEST v.2.0.

Table 4 proportion of examinee used

			7				8				10	
			(2/5)		(3/4)		(3/5)		(4/4)		(4/6)	
			a	b	a	b	a	b	a	b	a	b
Type 1 error	1000	Mean	0.789	0.953	0.742	0.997	0.804	0.993	0.730	1.000	0.851	0.999
		SD	0.086	0.084	0.081	0.018	0.064	0.022	0.081	0.001	0.048	0.006
	1500	Mean	0.742	0.997	0.725	0.998	0.802	0.997	0.725	1.000	0.837	0.996
		SD	0.081	0.018	0.081	0.012	0.064	0.017	0.084	0.000	0.071	0.013
	2000	Mean	0.805	0.976	0.727	0.999	0.803	0.996	0.724	1.000	0.848	0.998
		SD	0.071	0.053	0.082	0.006	0.064	0.020	0.085	0.000	0.078	0.011
	4000	Mean	0.808	0.987	0.728	1.000	0.803	0.997	0.726	1.000	0.850	1.000
		SD	0.071	0.044	0.080	0.001	0.063	0.018	0.082	0.000	0.066	0.002
Power	1000	Mean	0.799	0.992	0.729	1.000	0.800	0.999	0.726	1.000	0.840	1.000
		SD	0.059	0.029	0.068	0.000	0.071	0.007	0.081	0.002	0.064	0.003
	1500	Mean	0.797	0.995	0.727	1.000	0.799	1.000	0.728	1.000	0.841	1.000
		SD	0.058	0.021	0.066	0.000	0.070	0.001	0.080	0.000	0.064	0.000
	2000	Mean	0.797	0.999	0.726	1.000	0.800	1.000	0.729	1.000	0.841	1.000
		SD	0.058	0.009	0.066	0.000	0.070	0.000	0.079	0.000	0.063	0.000
	4000	Mean	0.798	1.000	0.747	1.000	0.803	1.000	0.747	1.000	0.864	1.000
		SD	0.058	0.003	0.084	0.000	0.070	0.000	0.079	0.000	0.058	0.000

a: DIMTEST v.2.0, b: DIMTEST v.2.1,

Table 5 shows the number of rejections out of 100 trials with 1 dimension for very short tests comparing DIMTEST v.2.0 with a modified version of DIMTEST v.2.0 that does not automatically eliminate examinees who get all right or all wrong on PT (DIMTEST v.2.1). The numbers of type 1 error with DIMTEST v.2.0 are clearly larger than those with the alternative DIMTEST with all examinees. However, the number of rejection rates is still much greater than the desired 0.05.

Table 5 Number of rejections out of 100 trials with 1 dimension (Type I error)

			No. Items (AT/PT)				
			7		8		10
			(2/5)	(3/4)	(3/5)	(4/4)	(4/6)
sample size	1000	a	13	18	13	15	10
		b	7	11	10	13	10
	1500	a	18	27	16	22	14
		b	11	21	6	17	13
	2000	a	21	31	23	29	20
		b	15	29	16	27	10
	4000	a	34	56	36	51	31
		b	27	50	28	47	26

a: DIMTEST v.2.0, b: DIMTEST v.2.1,

Table 6 shows a power study with short tests. Generally, the power is generally good enough, except for the shortest test. Still given the inflated Type 1 error rates above, these results must be viewed with caution.

Table 6 Number of rejections out of 100 trials with 2 dimensions (Power test)

			No. Items (AT/PT)				
			7		8		10
			(2/5)	(3/4)	(3/5)	(4/4)	(4/6)
sample size	1000	a	71	92	88	88	98
		b	67	87	84	93	87
	1500	a	79	97	100	99	99
		b	73	92	95	97	95
	2000	a	85	100	100	100	99
		b	83	97	94	97	98
	4000	a	95	98	100	98	100
		b	87	91	96	96	97

a: DIMTEST v.2.0, b: DIMTEST v.2.1,

DIMTEST with total scores vs. DIMTEST with Item response vectors

Four different DIMTEST are evaluated: DIMTEST v.2.0, DIMTEST v.2.1, and two DIMTEST using the examinee’s clusters based on item response vectors as PT cells with two different proximity matrix (hamming distance and p-weighted hamming distance, respectively DIMTEST_HMD and DIMTEST_PHMD). For this evaluation, we used 5 different lengths of test based on one sample size, N=1000 (more sample size are planned for a follow-up study.)

Table 7 Number of rejections out of 100 trials with 1 dimension

		MC=30				MC=1			
no.of items(AT/PT)		a	b	c	d	a	b	c	d
7	(2/5)	11	10	21	18	12	6	7	8
8	(2/6)	7	6	10	11	9	5	8	6
	(3/5)	8	3	29	26	8	3	19	19
10	(4/6)	5	3	51	48	8	3	30	32
15	(7/8)	3	6	57	64	2	2	36	32
	(4/11)	3	2	22	17	2	0	9	17
	(4/14)	2	1	11	15	1	0	7	11
18	(6/12)	4	3	34	47	2	2	18	26
	(8/10)	7	7	46	55	3	3	31	35
Total		62	50	457	480	62	42	278	310

a: DIMTEST v.2.0, b: DIMTEST v.2.1, c: DIMTEST_HMD, d: DIMTEST_PHMD

Table 7 shows number of rejections out of 100 trials with 1 dimension with four different DIMTESTs'. The numbers of rejections of DIMTEST v.2.1 are generally smaller that of DIMTEST v.2.0, regardless of the number of MC simulation. MC simulation is used to generate unidimensional simulated data sets. As the numbers of MC simulation are increased, the variance of the DIMTEST statistic based on unidimensional simulated data sets is reduced. If there is bias in the DIMTEST statistics, this increases possibility of Type 1 error rates. Also, this result is very similar to the previous results of Type 1 error (Table 5.) However MC=100 were used for the previous DIMTEST analyses. Thus, the numbers of Type 1 error on Table 7 are smaller than that of Table 5. The rejection rates of DIMTEST with item response vectors were even larger than the rejection rates from DIMTEST v.2.0. We are not sure why the vector approach results yielded lager Type 1 error rates.

Table 8 shows power comparisons with four different DIMTEST. As shown in Table 8, powers are overall good even with DIMTES with item response vectors. However, still given the inflated Type 1 error rates, the results must be viewed with caution.

Table 8 Number of rejections out of 100 trials with 2 dimension

no.of items	(AT/PT)	a	b	c	d
7	(2/5)	68	59	63	64
8	(2/6)	70	45	50	49
	(3/5)	90	83	87	88
10	(4/6)	94	89	95	94
15	(7/8)	100	90	99	100
	(4/11)	98	86	97	96
	(4/14)	97	95	94	96
18	(6/12)	100	99	99	99
	(8/10)	100	94	100	100
Total		1014	923	984	986

a: DIMTEST v.2.0, b: DIMTEST v.2.1, c: DIMTEST_HMD, d: DIMTEST_PHMD

Table 9 shows the proportion of the number of used examinees. The range of the proportion of examinee used with DIMTEST v.2.0, DIMTEST v.2.1, DIMTEST_HMD and DIMTEST_PHD, are respectively 0.794-0.969, 0.96-1.00, 0.871-0.984, and 0.873-0.981. DIMTEST_NHD and DIMTEST_PHD used 95% examinees (except in the case of 2 AT items and 6 PT items), and DIMTEST v.2.1 used 98% examinees regardless of test lengths. However, the proportion of the number of examinee used from DIMTEST v.2.0 is still sensitive to test length.

Table 9 Proportion of examinee used

no. of items (AT/PT)			Type 1 error				Power			
			a	b	c	d	a	b	c	d
7	(2/5)	Mean	0.803	0.979	0.920	0.919	0.794	0.988	0.944	0.946
		SD	0.071	0.050	0.076	0.079	0.062	0.042	0.066	0.067
8	(2/6)	Mean	0.833	0.960	0.871	0.873	0.857	0.988	0.910	0.912
		SD	0.078	0.071	0.092	0.101	0.060	0.039	0.063	0.061
	(3/5)	Mean	0.806	0.993	0.968	0.966	0.801	0.998	0.981	0.981
		SD	0.066	0.024	0.038	0.042	0.067	0.012	0.022	0.022
10	(4/6)	Mean	0.837	0.998	0.962	0.963	0.852	1.000	0.968	0.970
		SD	0.061	0.011	0.018	0.019	0.062	0.002	0.012	0.013
15	(7/8)	Mean	0.904	0.999	0.978	0.978	0.907	1.000	0.983	0.981
		SD	0.047	0.005	0.011	0.012	0.041	0.001	0.009	0.012
	(4/11)	Mean	0.936	0.985	0.965	0.961	0.954	0.998	0.975	0.971
		SD	0.038	0.034	0.022	0.021	0.023	0.005	0.011	0.012
	(4/14)	Mean	0.950	0.980	0.963	0.961	0.969	0.995	0.975	0.970
		SD	0.032	0.029	0.018	0.016	0.017	0.009	0.012	0.014
18	(6/12)	Mean	0.953	0.994	0.975	0.971	0.959	0.999	0.982	0.978
		SD	0.026	0.012	0.018	0.012	0.020	0.002	0.008	0.009
	(8/10)	Mean	0.938	0.997	0.980	0.976	0.942	0.999	0.984	0.979
		SD	0.029	0.010	0.010	0.012	0.028	0.001	0.008	0.011

a: DIMTEST v.2.0, b: DIMTEST v.2.1, c: DIMTEST_HMD, d: DIMTEST_PHMD

Table 10 shows the number of clusters and usable clusters which used for PT cells. The number of usable clusters became as expected, much larger than the numbers of PT items.

Table 10 Number of clusters and usable clusters

no.of items (AT/PT)		no.of possible cells	program	Type 1 error				power			
				no. of cluster		no. of usable clusters		no. of clusters		no. of usable clusters	
				Mean	SD	Mean	SD	Mean	SD	Mean	SD
7	(2/5)	32	c	30.03	1.59	22.52	4.28	29.74	1.73	22.56	4.05
			d	30.03	1.59	22.56	4.26	29.74	1.73	22.66	4.04
8	(2/6)	64	c	49.50	3.59	28.34	4.75	49.89	3.02	30.93	4.56
			d	49.33	3.55	29.04	4.99	49.64	2.97	31.41	4.54
	(3/5)	32	c	29.87	1.85	24.42	3.79	29.98	1.51	25.34	3.67
			d	29.87	1.85	24.51	3.67	29.98	1.51	25.40	3.66
10	(4/6)	64	c	49.79	3.04	35.02	3.84	49.01	2.56	35.35	3.79
			d	49.46	3.02	35.42	3.99	48.88	2.57	35.85	4.03
15	(7/8)	256	c	46.52	1.91	37.67	3.04	46.94	3.40	38.70	2.89
			d	48.92	3.61	38.84	3.56	49.97	5.04	39.47	3.14
	(4/11)	2048	c	47.72	2.18	37.08	3.23	50.04	3.28	38.30	3.30
			d	49.77	3.73	35.75	3.33	49.80	3.27	38.66	3.38
	(4/14)	16384	c	47.21	2.09	35.53	3.25	47.36	1.98	38.39	2.49

			d	48.90	2.92	35.05	3.26	49.37	3.05	37.00	2.81
18	(6/12)	4096	c	47.61	2.46	38.17	3.10	47.22	2.13	39.53	2.37
			d	49.23	3.32	36.69	3.22	48.88	3.42	38.16	2.47
	(8/10)	1024	c	47.58	2.05	39.10	2.87	47.74	2.31	40.02	2.64
			d	49.87	4.04	38.02	3.19	50.39	4.89	38.94	2.92
Total			c	45.06	7.88	34.04	6.79	44.90	7.75	35.14	6.72
			d	45.89	8.33	33.94	6.73	45.79	8.30	34.94	6.60
			Total	45.48	8.12	33.99	6.76	45.34	8.04	35.04	6.66

Discussion

DIMTEST with item response vectors had a higher rejection rates than the original DIMTEST. However, we found DIMTEST v.2.1 (that used 98% of examinee) gave improved results compared to DIMTEST v.2.0, especially with Type 1 error rates.

Even though our new alternative DIMTEST with item response vectors didn’t show any improvement in comparison to DIMTEST v.2.0 and DIMTEST v.2.1 with short test, we still believe that the procedure can be modified to obtain some advantages from this alternative DIMTEST for short tests, but much further work is needed. Based on this research result, we are going to look for the reasons why DIMTEST with item response vectors didn’t work.

References

Nandakumar, R., Yu, F, Li, H, and Stout, W. (1998), Assessing unidimensionality of polytomous data, *Applies Psychological measurement*, 22 , 99-115.

Roussos, L. A. (1992). HAC: Hierarchical Agglomerative Clustering Computer Program Users Manual.

Roussos, L. A., Sout, W. F., & Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement. 35*(1), 1-30.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika. 55, 293-326*

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.