

Evaluating DETECT Classification Accuracy and Consistency when Data Display Complex Structure

Mark J. Gierl
Jacqueline P. Leighton
Xuan Tan

Centre for Research in Applied Measurement and Evaluation
University of Alberta



Paper Presented at the 2005 Annual Meeting for the
Canadian Society for the Study of Education
London, Ontario

May 28-31, 2005

INTRODUCTION

- DETECT, the acronym for *Dimensionality Evaluation To Enumerate Contributing Traits*, is an innovative and relatively new nonparametric dimensionality assessment procedure
- It yields three different types of information that can be used to link substantive and statistical dimensionality analyses—(1) DETECT identifies the total number of dominant dimensions underlying student performance on a set of test items; (2) it estimates effect sizes to describe the amount of multidimensionality in a set of test items as well as the nature of the latent structure for these items and; (3) it specifies which single dimension is measured best by each test item
- To achieve these outcomes, DETECT identifies mutually exclusive, dimensionally homogeneous clusters of items using a genetic algorithm (Kim, 1994; Zhang & Stout, 1999)

INTRODUCTION

- The DETECT index is created by computing all item covariances after conditioning on the examinees' score using the remaining items; that is,

$$D(P) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq N} [Cov(X_i, X_j | \Theta_{TT})]$$

where n is the number of dichotomous items on a test, P denotes the partitioning of n items into k clusters, Θ_{TT} is the test composite, X_i and X_j are scores on items i and j

- When the data are unidimensional, clusters of items will be found that are not particularly homogeneous—in this case, the conditional covariance will be positive for some pairs of items and negative for other pairs of items resulting in a $D(P)$ index that is close to zero
- If, however, the data are multidimensional, then clusters of items will be found that have positive within-cluster conditional covariances and negative between-cluster conditional covariances, resulting in a $D(P)$ index that is greater than zero

INTRODUCTION

- To determine if the partition P , which produced $D(P)$ is, in fact, the correct partition to produce a simple structure solution, the following ratio is also computed:

$$r_{Max} = \frac{D(P)}{\bar{D}(P)}$$

- That is, r_{MAX} is an index representing how well the data approximates simple structure by comparing the maximum value of the partition to the average value of the conditional covariance across all item combinations
-

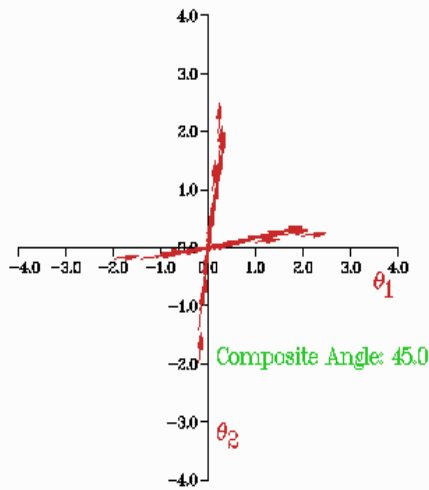
- When DETECT was evaluated by Zhang and Stout (1999) in a series of real and simulated data studies, the results were impressive as the number of correct runs ranged from 95 to 100% across their study conditions
- However, all analyses were limited to conditions where the data displayed approximate simple structure (i.e., $r_{MAX} \geq 0.80$)

INTRODUCTION

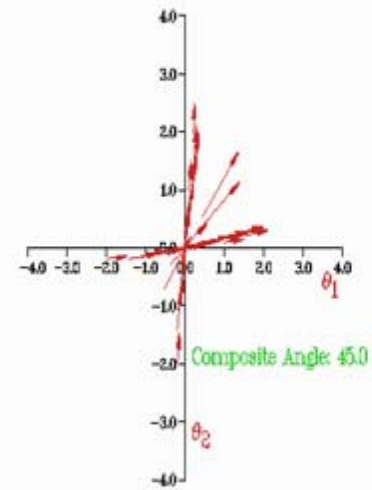
- In many testing situations, however, the data do display simple structure—rather, the data from many educational and psychological tests display a complex multidimensional structure, meaning the items on these tests measure two or more dimensions, but to differing degrees, in the multidimensional space
- No studies have been conducted to evaluate the performance of DETECT when the data display different degrees of complex structure (i.e., $r_{MAX} < 0.80$)
- Thus, the purpose of the current study was to evaluate DETECT item classification consistency and accuracy when the data display different degrees of complex structure in a two-dimensional space

METHODS AND DESIGN

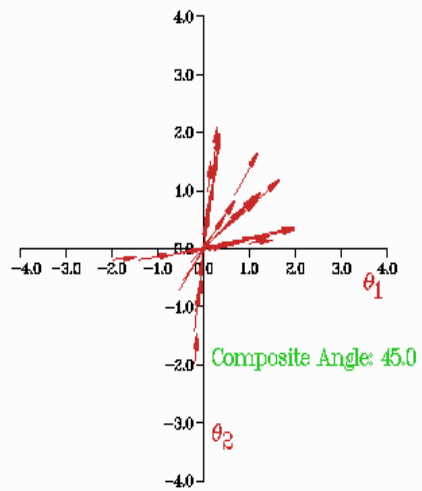
- Examinee response vectors to a 40-item tests were simulated for this study—the examinee ability estimates were generated from a bivariate normal distribution with a mean of zero and a standard deviation of one
- To generate the data, realistic item parameters were selected so our results would have some generalizability to different testing situations—the item parameters contained a range of a1-, a2-, and d-parameters, typical of what can found on multidimensional tests
- The items were also generated so they would measure both approximate simple and complex structure



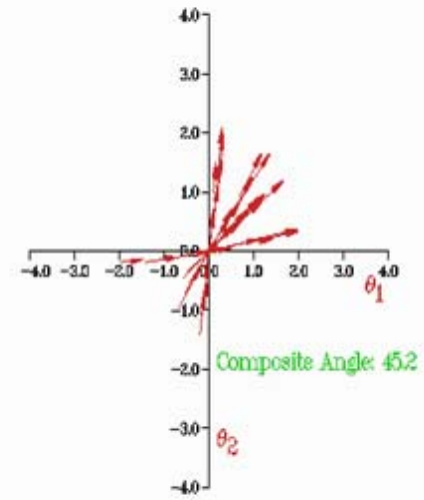
Condition #1: Simple Structure



Condition #2: Complex 10%



Condition #3: Complex 30%



Condition #4: Complex 50%

METHODS AND DESIGN

- Two factors were expected to influence item classification consistency and accuracy: The correlation between dimensions and the sample size—the correlation between dimensions was manipulated to range from no correlation to a high correlation (i.e., 0.0, 0.3, 0.6, or 0.9) while sample size was manipulated to range from small to large (i.e., 500, 1000, and 1500)
- The primary dependent variable measured classification accuracy by comparing item classification between the simulated data and true item parameters—that is, we compared the dimension each item was assigned to by DETECT with the actual dimension the item was intended to measure
- The mean and standard deviation for each correct match over 100 replication was computed to evaluate classification accuracy

RESULTS

Table 1. Overall Item Classification Accuracy

Correlation	Sample	Structure			
		Simple	Complex 10%	Complex 30%	Complex 50%
0.00	500	1.00 (0.01)	0.99 (0.01)	0.98 (0.02)	0.95 (0.03)
	1000	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.99 (0.01)
	1500	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)
0.30	500	1.00 (0.00)	0.99 (0.02)	0.96 (0.03)	0.91 (0.05)
	1000	1.00 (0.01)	1.00 (0.01)	0.99 (0.02)	0.97 (0.03)
	1500	1.00 (0.00)	1.00 (0.01)	0.99 (0.01)	0.99 (0.02)
0.60	500	0.99 (0.01)	0.98 (0.02)	0.92 (0.04)	0.84 (0.07)
	1000	1.00 (0.01)	0.99 (0.01)	0.96 (0.03)	0.90 (0.05)
	1500	1.00 (0.01)	0.99 (0.02)	0.97 (0.02)	0.94 (0.04)
0.90	500	0.69 (0.12)	0.67 (0.13)	0.62 (0.11)	0.57 (0.09)
	1000	0.87 (0.12)	0.81 (0.14)	0.77 (0.09)	0.60 (0.10)
	1500	0.91 (0.14)	0.85 (0.14)	0.81 (0.09)	0.63 (0.10)

RESULTS

Table 2. Item Classification Accuracy For Dimensions 1 and 2

Correlation	Sample	Structure							
		Simple		Complex 10%		Complex 30%		Complex 50%	
		1	2	1	2	1	2	1	2
0.00	500	1.00 (0.01)	1.00 (0.00)	0.99 (0.02)	0.99 (0.02)	0.98 (0.03)	0.98 (0.03)	0.95 (0.05)	0.96 (0.04)
	1000	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.02)	0.99 (0.02)	0.99 (0.02)
	1500	1.00 (0.01)	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	1.00 (0.02)	1.00 (0.01)
0.30	500	1.00 (0.01)	1.00 (0.00)	0.99 (0.02)	0.99 (0.02)	0.96 (0.05)	0.95 (0.04)	0.91 (0.06)	0.91 (0.07)
	1000	1.00 (0.01)	1.00 (0.00)	0.99 (0.02)	1.00 (0.01)	0.99 (0.02)	0.98 (0.03)	0.98 (0.04)	0.97 (0.04)
	1500	1.00 (0.00)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.99 (0.02)	0.99 (0.02)	0.99 (0.03)	0.99 (0.03)
0.60	500	0.99 (0.02)	1.00 (0.01)	0.97 (0.03)	0.98 (0.03)	0.93 (0.06)	0.91 (0.06)	0.84 (0.09)	0.83 (0.09)
	1000	1.00 (0.01)	1.00 (0.01)	0.99 (0.02)	0.99 (0.02)	0.97 (0.04)	0.95 (0.05)	0.93 (0.07)	0.88 (0.09)
	1500	1.00 (0.01)	1.00 (0.00)	0.99 (0.03)	0.99 (0.02)	0.98 (0.03)	0.97 (0.04)	0.94 (0.07)	0.94 (0.06)
0.90	500	0.70 (0.12)	0.68 (0.14)	0.69 (0.13)	0.66 (0.16)	0.62 (0.12)	0.62 (0.13)	0.58 (0.12)	0.57 (0.12)
	1000	0.88 (0.11)	0.87 (0.14)	0.82 (0.14)	0.80 (0.16)	0.81 (0.13)	0.72 (0.10)	0.62 (0.14)	0.58 (0.13)
	1500	0.92 (0.12)	0.91 (0.18)	0.86 (0.12)	0.84 (0.18)	0.84 (0.13)	0.77 (0.12)	0.64 (0.19)	0.62 (0.16)

RESULTS

Table 3. Item Classification Accuracy for Complex Structure Items

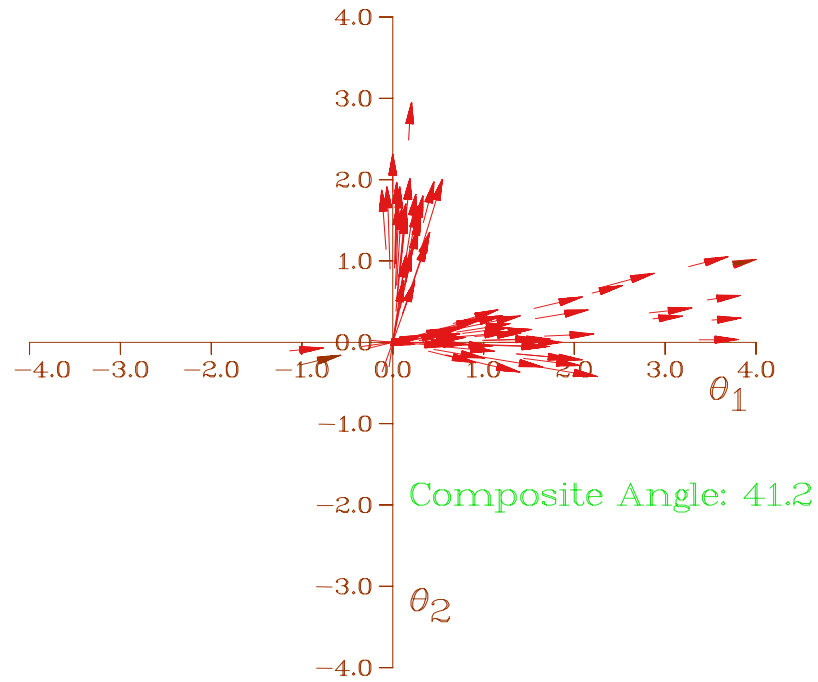
Correlation	Sample	Complex Structure		
		10%	30%	50%
0.00	500	0.93 (0.14)	0.92 (0.08)	0.91 (0.06)
	1000	0.98 (0.07)	0.99 (0.03)	0.98 (0.02)
	1500	1.00 (0.02)	1.00 (0.02)	0.99 (0.02)
0.30	500	0.89 (0.16)	0.86 (0.11)	0.82 (0.09)
	1000	0.96 (0.09)	0.96 (0.06)	0.94 (0.06)
	1500	0.99 (0.05)	0.98 (0.04)	0.98 (0.04)
0.60	500	0.81 (0.20)	0.74 (0.12)	0.70 (0.11)
	1000	0.88 (0.15)	0.86 (0.10)	0.81 (0.11)
	1500	0.90 (0.16)	0.91 (0.08)	0.90 (0.09)
0.90	500	0.57 (0.23)	0.47 (0.16)	0.52 (0.10)
	1000	0.60 (0.22)	0.58 (0.13)	0.50 (0.12)
	1500	0.58 (0.19)	0.60 (0.14)	0.52 (0.11)

CONCLUSIONS

- The purpose of this study was to evaluate DETECT item classification consistency and accuracy when the data display different degrees of complex structure in a two-dimensional space because all empirical studies evaluating the properties of DETECT, to-date, have focused simple structure data
- The overall classification rates were high (i.e., >90%) for almost all dimensional structures when the correlation was 0.00, 0.30, and 0.60 across sample sizes; when the correlation was 0.90, the accuracy rates fell below 90% for almost all study conditions (similar results were found when classification consistency was assessed separately for Dimensions 1 and 2)
- When item classification accuracy was evaluated for the complex structure items, the rates met or exceeded 90% only when the correlation was 0.00 for all sample size conditions; with a correlation of 0.30 and a sample size of either 1000 or 1500, accuracy exceeded 90%; with a correlation of 0.60 and a sample size of 1500, accuracy exceeded 90%; with a correlation of 0.90, accuracy rates were below 90% for all samples

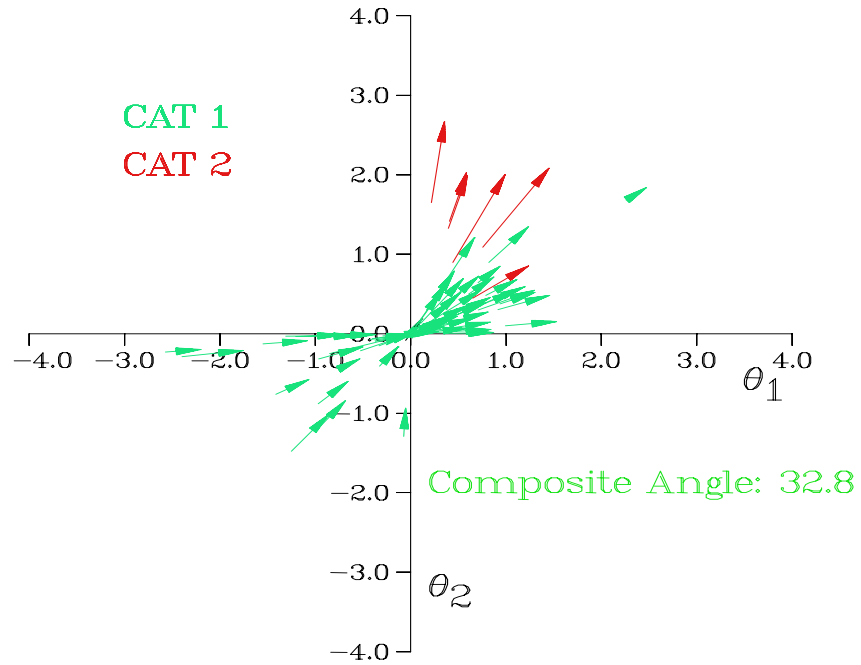
IMPLICATIONS FOR PRACTICE

- Given the conditions evaluated in this study, we conclude that DETECT can adequately classify items—meaning the classification rates are equal to or greater than 90%—in a two-dimensional space for some complex structures
- More specifically, the results from our study indicate that DETECT can accurately classify items according to their true underlying dimensional structure when as many as 30% display complex structure, if the correlation is $r \leq 0.60$ and $n \geq 500$; if 50% of the items display complex structure, then 1000 examinees or more are required
- But when $r = 0.90$ DETECT does not work well with any dimensional structure regardless of sample size



LSAT Logical and Analytic Reasoning Items: Simple Structure

[Back](#)



Alberta Learning Language Arts 9: Complex Structure