Cluster Analysis for Cognitive Diagnosis: An Application to the 2001 PIRLS Reading Assessment

Chia-Yi Chiu, Rutgers, The State University of New Jersey, chia-yi.chiu@gse.rutgers.edu
Minhee Seo, University of North Carolina at Greensboro, minseo@uiuc.edu
Jeff Douglas, University of Illinois at Urbana-Champaign, jeffdoug@illinois.edu

Abstract

Latent class models for cognitive diagnosis often begin with specification of a matrix that indicates which attributes or skills are needed for each item. Then by imposing restrictions that take this into account, along with a theory governing how subjects interact with items, parametric formulations of item response functions are derived and fitted. Cluster analysis provides an alternative approach that does not require specifying an item response model, but does require an item-by-attribute matrix. After summarizing the data with a particular vector of sum-scores, K-means cluster analysis or hierarchical agglomerative cluster analysis can be applied with the purpose of clustering subjects who possess the same skills. An application study to the 2001 PIRLS reading data is conducted to illustrate how the methods can be implemented in practice.

Keywords: *cluster analysis, cognitive diagnosis, Q matrix.*

Introduction

Most existing large-scale assessments, including PIRLS, only report students' overall performances. Along with the increasing reliance on this type of assessments, there is pressure to make assessments more diagnostically informative about required skills and cognitive processes they measure in students (Leighton & Gierl, 2007). In the cognitive diagnosis context, classification based on students' mastery or non-mastery of each attribute in a set of attributes is desired. When a cognitive diagnostic model (CDM) is specified, classification can be done by fitting the model, and estimating parameters through likelihood functions. However, accessing complex cognitive diagnostic models requires sophisticated software. Few of these programs are in the public domain, and the existing programs tend to apply EM algorithm, or Markov chain Monte Carlo, for which convergence can be slow and is never certain. Exploratory cluster analysis, on the other hand, provides an alternative that does not require specifying a cognitive diagnosis model, even though a specific model is assumed. This requires clustering on a properly chosen summary score of the data, as well as some assumptions provided by experts to identify item-to-skill information used in the latent class models for cognitive diagnosis. One advantage is that no further model assumptions are needed. By applying the cluster analysis, users can run familiar and widely available software to conduct classification study, and depending on the method that is used, computer run time can be very short. Additionally, considering the goal of measuring progress of students' learning, it is important for PIRLS not only to report overall assessment outcomes, but also to provide more fine-grained information on skill mastery. It is thus of interest to apply the newly developed clustering theory to the PIRLS data and see how a complete cognitive diagnostic analysis can go from here.

Restricted Latent Class Models for Cognitive Diagnosis

Specialized latent class models for cognitive diagnosis are formulated under assumptions on which attributes are needed for which items, and how the attributes are utilized to construct a response. Let $\boldsymbol{\alpha}$ be a K-dimensional vector for which the k^{th} entry α_k, indicates whether or not a subject possesses the k^{th} attribute or skill, for $k=1,2,\ldots,K$. An attribute might refer to a clearly defined skill in some applications, or a more abstract psychological construct in another. All restricted latent class models for cognitive diagnosis that we consider require a $J\times K$ matrix **Q**, referred to as a Q-matrix (Tatsuoka, 1985), with the (j, k) entry q_{jk} denoting whether or not the j^{th} item requires the k^{th} attribute. The vector $\boldsymbol{\alpha}$ can take 2^K distinct values. These values index the 2^K latent classes in such models. What distinguishes models from one another are the assumptions that dictate how attributes are utilized to construct responses. In this study, we give an example used for the spplication. A recent and thorough review of latent class models for cognitive diagnosis can be found in Rupp and Templin (2007).

DINA Model

The DINA (Deterministic Input, Noisy Output "AND" gate) model (Junker and Sijtsma, 2001) is a conjunctive model. It extends the work of Macready and Dayton (1977), which considers a two-class version of it for assessing mastery of a skill. The item response function of the DINA model is,

$$P(Y_{ij}=1|\alpha_i)=(1-s_j)^{\eta_{ii}}\,g_j^{(1-\eta_{ii})}\ ,$$

where for all i, $s_j=P(Y_{ij}=0|\eta_{ij}=1)$ and $g_j=P(Y_{ij}=1|\eta_{ij}=0)$ are the probabilities of slipping and guessing, respectively, for the j^{th} item, and η_{ij} is the ideal response which connects the attribute pattern possessed by a subject and the elements of **Q** in the following way:

$$\eta_{ii}=\prod_{k=1}^{K}\alpha_{ik}^{q_{ik}}\ .$$

The variable η_{ij} indicates whether the subject possesses all the attributes needed for answering the particular item. Therefore, the DINA model is characterized by its conjunctive feature that the probability of answering an item correctly will severely drop if any of the required attributes are not mastered or possessed. Fitting the DINA model can be done with the EM algorithm (Haertel, 1989), or by use of Markov chain Monte Carlo (de la Torre and Douglas, 2004; Tatsuoka, 2002). Templin et al. (2007) discuss how to fit cognitive diagnosis models, including the DINA model as well as the remaining models in this section, using the software Mplus (Muthén and Muthén, 2006).

Cluster Analysis for Cognitive Diagnosis

The latent variable models discussed in the previous section all require sophisticated software for fitting, either with the EM algorithm or by Markov chain Monte Carlo. In this part, we will describe how one can construct appropriate sum-scores that can be used to cluster examinees into the correct latent classes. We first introduce the sum-score statistic, then discuss two common methods of cluster analysis that can be used to arrive at the classifications.

The method begins by first constructing a vector of sum-scores. For the i^{th} examinee, the k^{th} entry of the variable $\boldsymbol{W}_i=(W_{i1},W_{i2},\ldots,W_{iK})'$ is defined as,

$$W_{ik} = \sum_{j=1}^{J} Y_{ij} q_{ik} \ .$$

Therefore, the entries of $\boldsymbol{W}$ are sum-scores for items corresponding to each attribute. The vector $\boldsymbol{W}$ is then taken as the input to a user-chosen method of cluster analysis, with a fixed number 2^K clusters. The methods of cluster analyses we have investigated for this application are K-means (MacQueen, 1967), and hierarchical agglomerative cluster analysis, which will be referred to as HACA. Supporting by the asymptotic theory, to be proposed later, HACA with complete linkage will be able to classify data, under the DINA structure, to correct groups.

K-means

K-means cluster analysis is a widely used exploratory technique to cluster subjects based on a vector of data. The K-means algorithm is mostly about estimating the cluster centers with the number of clusters being pre-determined. Once the centers are decided, data are sent to the closest cluster. Specifically, consider a data matrix with N subjects and K variables, where the row entries are $\boldsymbol{w}_1, \boldsymbol{w}_2,\ldots,\boldsymbol{w}_N$, and the goal is to cluster the N subjects into M clusters, where $M=2^K$, based on the observations of the K variables. Then taking Euclidean distance as an instance, the K-means method assigns data point $\boldsymbol{w}_i$ to the m^{th} cluster if

$$m = \arg \min_{u\in\{1,\ldots,M\}} \left\| w_i - \hat{c}_u \right\|^2 \qquad (1)$$

Where $\boldsymbol{c}_u$ is the estimated center of the u^{th} cluster and is the average of the data in the cluster. The iteration to find out the final solution is as follows:

1. Choose *M* initial *K*-dimensional cluster centers.
2. Assign data points to clusters by the criterion (1).
3. Obtain the updated cluster center by calculating the average of assigned observations.
4. Repeat 2 and 3 until no observation can be relocated.

One shortcoming about this K-means method is that the final solutions rely heavily on the selection of the initial values. Having poor starting values can result in convergence to local optima (Steinley, 2003), and can yield solutions that are much poorer than the global solution. Many methods of initializing starting values for the K-means algorithm have been proposed. MacQueen (1967) suggested selecting the first *M* data points as the initial starting values. As we can see, this method may cause problems when data are ordered in some particular way. A modification of this method is to, instead of selecting the first *M* data points, randomly select *M* data points over the data set. Also regarded as an adjusted method of random selection, Forgy (1965) proposed an initialization method of first randomly selecting *M* data points as seeds, assigning remaining data points to the cluster with the closest seed, computing the means of the clusters, and then taking these means as starting values for the K-means algorithm. For other methods, Steinley (2006) had more thorough reviews.

HACA (Hierarchical Agglomerative Cluster Analysis)

In contrast to partitioning data into exclusive clusters like K-means does, hierarchical clustering forms a dynamic tree structure of clustering in which not only are the distances between data points taken into account, but also the distances between clusters are considered. Compared with K-means, HACA is much simpler computationally, and does not require selecting initial values. HACA begins by defining a matrix of distances for all pairs of distinct observations, say Euclidean distance in our case. Then each subject begins as its own cluster. The next step is to cluster the two objects for which the distance between them is smallest. At each step thereafter, two clusters are fused by adjoining two of the existing clusters. Defining these distances between clusters is what distinguishes different methods of linking clusters, which comprise the variations of HACA. Clusters are combined by using one of the following linkages to minimize the distance between the clusters that are joined in each step until the process is stopped at a fixed number of clusters or until only one cluster containing all of the objects remain. In our application, the process is stopped at the point where there are 2^K clusters.

Some common linkages are introduced in the following. The first is the complete linkage, in which the distance between clusters is the maximum distance between two data points from the two targets clusters. Complete linkage clustering tends to produce homogeneous clusters. Single linkage, on the other hand, defines distance according to the minimum distance resulting from taking a point from each cluster. Single linkage tends to produce long, stringy clusters and non-convex shapes, which is known as the chaining effect. Findings (Chiu, Douglas, & Li, 2008; Chiu, Douglas, 2008) showed that this method performs poorly in similar applications, we would not employ this linkage in this study. Instead of taking the two extreme distances into consideration, average linkage clustering uses the mean of distances between the data points in two different clusters as a measure. In addition, Ward (1963) proposed a general hierarchical clustering method in which clusters are chosen to merge so that the updated within cluster sum of square errors are minimized. This Ward's linkage tends to produce nearly equal sized clusters which are convex and compact, and thus suffers from outliers (Milligan, 1980).

Methodology

Asymptotic Theory for Classification

The asymptotic theory for classification under a given cognitive diagnostic model using cluster analysis has been developed and mathematically proved by Chiu, Douglas and Li (submitted, 2008). If one would like to successfully borrow clustering techniques to classify data under some certain model, the key is to come up with an appropriate sample statistic as an input to access to the clustering algorithm. The sample statistic taken in this study is the sum score $\boldsymbol{W}$ introduced in the previous section. The asymptotic theory is supported by three lemmas which connect and function in the following way. The first lemma indicates how identifiability of attribute patterns is to be achieved. Assuming that data are generated according to the DINA model, the second lemma makes use of the previous lemma to show that the expectation of $\boldsymbol{W}$ is distinct for each of the 2^K attribute patterns. The final lemma relates to the behavior of HACA when all of the observations are close to their expected values, which is what takes place with high probability for long exams when the data are summarized by the vector $\boldsymbol{W}/J$. These preceding lemmas are now used in a proof of the consistency of clustering, along with a corollary to show that the theorem holds in the case of the DINA model. The theory is stated in the following. Assuming that there are N subjects, J items, and K attributes, the proposed sample statistic is a vector of sum-scores, $\boldsymbol{W} = (W_1, W_2, \cdots, W_K)$, where the k^{th} component, W_k is defined as:

$$W_k = \sum_{j=1}^{J} Y_j q_{jk} .$$

Thus, the components of vector $\boldsymbol{W}$ are just sum-scores for items corresponding to each attribute. Additionally define variable $\boldsymbol{V}=\boldsymbol{W}/J$, which is a re-scaled $\boldsymbol{W}$, and will give the same classification as $\boldsymbol{W}$ in real cases. The framework of the theory is as follows:

*Let G(i) and G(i') denote cluster assignments of subjects i and i', when using hierarchical agglomerative clustering technique with an appropriate linkage, through the statistic **V**. With some regularity conditions on cognitive diagnosis model, then as* $J \to \infty$*, and* $N/e^J \to 0$,

$$P[G(i)=G(i') \mid \boldsymbol{a}_i=\boldsymbol{a}_{i'}] \to 1 \quad \text{for all} \quad i \neq i'$$

and

$$P[G(i) \neq G(i') \mid \boldsymbol{a}_i \neq \boldsymbol{a}_{i'}] \to 1 \quad \text{for all} \quad i \neq i',$$

where $\boldsymbol{a}_i$ *is the attribute pattern for the* i^{th} *subject.*

This is the asymptotic theory expressed in a more understandable way. For more careful details, please refer to Chiu, Douglas, Li (submitted, 2008). Because of its asymptotic nature, in the real settings, we need a long test to ensure a more accurate classification. In the theory, the DINA model is assumed for some of the main results, though the only critical condition needed to generalize the consistency theorem to other models is that the mean of the vector $\boldsymbol{W}$ is distinct for different values of the binary latent vector $\boldsymbol{a}$.

Data Source

The 2001 PIRLS reading assessment data are used for this empirical application. The reading comprehension test consists of six sets of items, each set based on a different reading passage or topic. Each set includes 11 to 14 items and was administered to 25% of the examinees. Because the required skills are specified in terms of “purposes” and “process” for each item, this information is used to construct the corresponding Q-matrix. The following is an example:

[Take in Table 1 about here]

To apply the classification method, all examinees have to take same items, so examinees are grouped according to which booklets they took, and classified within the group. The distribution of examinees to booklets is as follows:

[Take in Table 2 about here]

It is known that higher number of attribute patterns would lower the accuracy of classification (Chiu, Douglas, & Li, submitted), and thus require more items to achieve interpretable results. As we can see from the table that for each group, no more than 20 items were taken. According to previous study (Chiu, Douglas, 2008), when K=4, tests shorter than 20 items would result in poor classifications. Therefore, only the skills constructed based on assessment purpose will be used in the study. For those constructed according to assessment process will not be included in this study. Furthermore, only binary items are taken into account and the study will proceed based on the following design:

- The sum-scores vector, ***W***, is used as an input to the clustering algorithms.
- Both K-means and HACA will be utilized to classify data.
- DINA model is fitted as a contrast, using EM algorithm for parameter estimation.
- Cluster size, within-cluster mean of ***W***, and within-cluster sum squares (WCSS) of ***W*** were used to evaluate classification for each method. ARI is also applied to indicate the agreement between methods.

Because there are quite many data sets to be analyzed using several methods, the long computer running time may be a concern. Therefore, from each group, we randomly drew 5000 examinees which are about half of the examinees in each group.

Analysis Results

Based on the booklets that the examinees took, the data were divided into 10 groups. However, as mentioned in lemma 1, for a test to identify all possible examinees' attribute patterns, all possible single skill items have to be included. There were only 6 sets of items covering both required skills and thus were used in the analysis. Each group of examinees was classified by applying cluster analysis techniques, including K-means, and HACA with complete, average, and Ward's linkages, as well as fitting the DINA model with EM algorithm as a contrast. For the DINA, clusters were labeled with the attribute pattern which maximized the posterior likelihood. However, attribute labels are not directly available for K-means and HACA, and this makes interpretation more difficult. Taking advantage of the feature of partial ordering among the 4 attribute patterns, the results from K-means and HACA were sorted along with the sum scores of Y, without assuming any model.

Table 3 shows the classification results for the data from Group B who took booklets 1 and 6.

[Take in Table 3 about here]

Although the true model is unknown, the clear patterns of mean ***W*** in the DINA-EM part indicates that examinees are adequately classified based on their attribute patterns. In the cluster analysis part, HACA with complete and average linkage also have recognizable mean ***W***, while HACA with Ward's linkage and K-means have some unrecognizable mean ***W*** patterns which is a sigh of not being able to separate data accordingly. The ARI table shows the result consistent to the former finding, in which HACA with complete and average linkages have higher ARI's with DINA-EM than the others.

Table 4 concludes the results for Group C who took booklets 1 and 8.

[Take in Table 4 about here]

Different from the results of Group B, the DINA-EM in this case does not fit the data well, based on the confounded mean ***W*** it shows. The classifications from HACA and K-means are not recognizable either, but provide more information than that from DINA-EM.

Table 5 shows the results for Group E who took booklets 2 and 5.

[Take in Table 5 about here]

Similar to those from Table 4, DINA is not considered as an adequate model for the data, with all mean ***W***'s locating at the same spot. Among the cluster analysis methods, K-means outperforms the others in which the mean ***W*** patterns are recognizable.

Table 6 shows the results for Group F who took booklets 3 and 7.

[Take in Table 6 about here]

In this case, HACA with average linkage performs better than the other methods in the sense that the data were better classified into the correct cluster. The other methods obviously did not separate data appropriately based on their attribute patterns.

Table 7, which shows the results of Group G who took booklets 4 and 6, has the similar results to those in Table 4, in which classification seems to be difficult for all methods.

[Take in Table 7 about here]

The difference is that the data in this case are more evenly assigned to clusters. If this is the true structure of the data, K-means or HACA with Ward's linkage would be considered to be more reliable methods because they tend to produce clusters with similar sizes.

For Table 8, examinees took booklets 4 and 8, and HACA with complete linkage produced more recognizable mean ***W*** than the other methods.

[Take in Table 8 about here]

Although DINA-EM has higher agreement with HACA with complete linkage than the other methods, the mean ***W*** from DINA-EM does not produce interpretable patterns for pointing out the underlying examinees' attribute patterns.

Conclusion and Implications

The classification theory is built up on the assumption of long tests. The application to the PIRLS data is a good opportunity to understand how robust the method is for short tests, but it also points out the fact that if it is the intension to use the proposed method to classify examinees for the cognitive diagnosis purpose, examinees need to take more items and the contents of the items should be designed to fit the needs.

The previous findings related to this cluster analysis method for classifying examinees under the cognitive diagnosis setting indicated that this method, depending on what type of structure the data set has, is more robust than fitting a wrong model (Chiu, Douglas, 2008). This is consistent with the results in the current study. When DINA-EM fits the data well, there would be at least a cluster analysis which performs equally well as the DINA-EM; while the DINA does not appear to be the true model, one could find some cluster analysis which is more reliable. There is more work on labeling which is critical and important for the K-means and HACA, but at this point, the feature of easy access and the convincing results from simulations and real applications make the theory an appreciable alternative for classification.

References

IEA. (2001). PIRLS 2001 User Guide for the international database. Retrieved Jan, 2007, from http://isc.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html

Chiu, C., & Douglas, J. (2008, July). *Cluster Analysis for Cognitive Diagnosis: A Robustness Study in Relation to Model Misspecification.* Paper presented at the annual meeting of the Psychometric Society, Durham, NH.

Chiu, C., Douglas, J. & Li, X. (2008) Cluster Analysis for Cognitive Diagnosis: Theory and Applications, under review.

de la Torre, J., & Douglas, J. A. (2004). Higher order latent trait models for cognitive diagnosis, *Psychometrika, 69,* 333-353.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics, 21*, 768-769.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*, 333-352.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological measurement*, *25*, 258-272.

Leighton, J. & Gierl, M. (2007) Why Cognitive Diagnostic Assessment? *Cognitive Diagnostic Assessment for Education* (pp. 3-18), Cambridge: Cambridge University Press.

MacQueen, J. (1967). Some methods of classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Bekeley Symposium on Mathematical Statistics and Probability* (pp. 281-207). Berkeley, CA: University of California Press.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *2*, 99-120.

Milligan, G. W. (1980). An examination of the effects of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325-342.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide*. (4th ed.). Los Angeles, CA: Muthén & Muthén.

Rupp, A. A., & Templin, J. L. (2007, April). Unique characteristics of cognitive diagnosis models. *The Annual Meeting of the National Council for Measurement in Education.* Chicago.

Steinley, D. (2003). Local optima in K-means clustering: What you don't know may hurt you. *Psychological Methods*, *8*, 294-304.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology, 59*, 1-34.

Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics, 10*, 55-73.

Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Applied Statistics (JRSS-C), 5*, 337-350.

Templin, J., Henson, R., & Douglas, J. (2007). General theory and estimation of cognitive diagnosis models: Using Mplus to derive model estimates. Unpublished manuscript.

Ward, J. H. (1963). Hierarchical Grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.

Tables

Table 1: An example of Q-matrix based on assessment purposes

Item	Purpose	Q-matrix	
40	Acquire and Use Information	1	0
41	Acquire and Use Information	1	0
42	Acquire and Use Information	1	0
43	Acquire and Use Information	1	0
44	Acquire and Use Information	1	0
45	Acquire and Use Information	1	0
46	Acquire and Use Information	1	0
47	Acquire and Use Information	1	0
48	Literary Experience	0	1
49	Literary Experience	0	1
50	Literary Experience	0	1
51	Literary Experience	0	1
52	Literary Experience	0	1
53	Literary Experience	0	1

Table 2: The distribution of examinees to booklets taken

			Examinees' Group ID											
Booklet	Required skill	# of dichot. items	A	B	C	D	E	F	G	H	I	J	K	L
1	AI[a]	4	x	x	x									
2	AI	6	x			x	x							
3	AI	8						x	x	x				
4	AI	3				x					x	x	x	x
5	LE[b]	6					x						x	
6	LE	7		x							x			
7	LE	5						x	x	x				
8	LE	7			x							x		x

[a]AI: Acquire and use information

[b]LE: Literary Experience

Table 3: Classification results for Group B who took booklets 1 and 6.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
903	2.3	4.2	4204	295	1.2	2.1	565
363	4.3	2.9	782	5	6.4	1.4	4
40	2.2	9	45	1193	3.5	4.8	4594
3694	5.7	8.1	11504	3507	5.8	8.2	10755
Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
1270	3	3.7	5981	880	2.5	3.1	3265
1277	4.3	8.1	2214	1261	5.5	6	2472
1290	6.2	6.9	2180	887	3.8	7.8	1653
1163	6.6	9.5	573	1972	6.3	9	2123

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1390	3	4	6776
(1 0)	97	6.2	4	109
(0 1)	132	2.7	8.4	131
(1 1)	3381	5.9	8.3	8793

ARI	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.74	0.7	0.3	0.39
Comp	*	1	0.75	0.21	0.33
Ave	*	*	1	0.26	0.35
Ward	*	*	*	1	0.34
K-mns	*	*	*	*	1

Table 4: Classification results for Group c who took booklets 1 and 8.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
436	2.7	2.5	1448	870	2.6	3.5	3762
878	2.7	5.8	2821	17	0.5	8.1	41
1127	5.7	6.9	2146	1512	5	6.9	3804
2559	5.8	10	6741	2601	5.7	10	7491
Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
590	2.6	2.8	2451	781	2.4	3.3	3186
1121	3.5	6.1	2847	1216	4	6.6	3415
1306	5.2	8.1	4090	1360	5.9	8.4	1873
1983	6.3	10.3	2906	1643	6.1	10.7	2505

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1540	5	7	8691
(1 0)	95	5.1	6.8	100
(0 1)	97	4.9	6.7	116
(1 1)	3268	5	7	10548

ARI	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.47	0.41	0.42	0.43
Comp	*	1	0.82	0.57	0.45
Ave	*	*	1	0.52	0.39
Ward	*	*	*	1	0.55
K-mns	*	*	*	*	1

Table 5: Classification results for Group E who took booklets 2 and 5.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
1481	2.7	3.3	6165	1360	3.2	2.2	4738
300	4.8	1.6	388	110	2.4	5.1	144
1366	5.9	5.7	3340	1955	5.1	6	6159
1853	5.9	8.6	5598	1575	6.4	8.8	3132
Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
1107	2.3	2.8	4180	1228	2.8	2.2	3653
894	5.1	3.1	2435	989	3.8	6.4	2016
1884	5.2	6.9	5045	903	5.7	4.5	2151
1115	6.7	9.1	1577	1880	6.4	8.5	3869

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1914	5	7	8463
(1 0)	62	5	7	87
(0 1)	78	5.1	7.2	100
(1 1)	2946	5	7	10981

ARI					
	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.42	0.37	0.38	0.4
Comp	*	1	0.52	0.42	0.46
Ave	*	*	1	0.46	0.59
Ward	*	*	*	1	0.4
K-mns	*	*	*	*	1

Table 6: Classification results for Group F who took booklets 3 and 7.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
847	3	2.8	2891	241	1.7	1.5	457
113	1.3	5.4	103	1333	3.2	4.7	4033
2255	5.1	6.2	5398	163	6.6	3.4	260
1785	8.2	6.1	3832	3263	7	6.3	9916
Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
690	2.7	2.6	1986	610	2.6	2.4	1948
765	2.9	5.6	1278	1010	3.2	5.6	1852
1558	5.4	5.8	2761	2064	6	5.9	3459
1987	8.1	6.3	3297	1316	8.7	6.4	1499

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1241	2.7	4.1	5337
(1 0)	9	6	5.1	4
(0 1)	753	2.7	6.9	1232
(1 1)	2997	5.5	8	7899

ARI					
	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.31	0.67	0.43	0.38
Comp	*	1	0.27	0.54	0.46
Ave	*	*	1	0.42	0.38
Ward	*	*	*	1	0.54
K-mns	*	*	*	*	1

Table 7: Classification results for Group G who took booklets 4 and 6.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
304	1.7	1.4	599	316	1.9	1.4	776
940	1.9	4.9	2883	779	1.9	4.3	1462
777	5	5.6	1675	484	4.8	4.9	658
2979	4.6	8.5	7991	3421	4.5	8.3	10540

Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
828	1.5	3.3	3251	785	2	2.6	2352
892	4.1	4.8	1773	942	2.3	6.5	2048
1612	3.5	7.9	3668	1060	4.7	6.1	1643
1668	5.5	8.9	1738	2213	5	9	3189

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1476	4.4	6.9	6491
(1 0)	143	4.6	6.9	190
(0 1)	171	4.5	7.1	208
(1 1)	3210	4.4	6.9	8243

ARI					
	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.54	0.58	0.31	0.42
Comp	*	1	0.74	0.34	0.53
Ave	*	*	1	0.41	0.39
Ward	*	*	*	1	0.41
K-mns	*	*	*	*	1

Table 8: Classification results for Group H who took booklets 4 and 8.

Cluster Analysis							
Complete				Average			
Size	mean W		WCSS	Size	mean W		WCSS
779	1.8	2.8	2702	32	3.2	0.6	28
586	2.1	7	1072	855	2.1	3.1	3033
589	4.1	5.6	950	1091	3.4	6.3	3215
3046	4.8	9.6	8921	3022	4.6	9.7	9254

Ward				K-means			
Size	mean W		WCSS	Size	mean W		WCSS
1117	2	3.5	4716	735	1.9	2.6	2298
1667	3.6	7.5	4668	1142	3.1	6.1	3098
659	5.4	8.6	316	1634	4.4	8.5	3320
1557	5	10.7	2317	1489	5.1	10.7	1992

DINA-EM				
Pattern	Size	mean W		WCSS
(0 0)	1397	4.4	6.8	7185
(1 0)	91	4.6	6.9	170
(0 1)	193	4.6	7	219
(1 1)	3319	4.4	6.9	11614

ARI					
	DINA	Comp	Ave	Ward	K-mns
DINA	1	0.7	0.55	0.28	0.33
Comp	*	1	0.72	0.28	0.48
Ave	*	*	1	0.34	0.46
Ward	*	*	*	1	0.58
K-mns	*	*	*	*	1